

CS – CODED BIAS Rubric

Elizabeth Breslin

DS 4002 – Fall 2022 – Professor Alonzi

Due: Dec 14

Submission Format: Link to gitbub repository (collab assignments) and hard copy (in class)

Individual Assignment

General Description: Submit to collab assignments a link to your case study repository and a hard copy of the README.md file to the head TA in class.

Preparatory Assignments – Read articles provided in the materials section for a background on facial recognition technologies. The Kaggle, messy and clean room identification model code that will be adapted to create the gender identification model.

Why am I doing this? We read and produce solutions to case studies to practice thinking like a data scientist. In this example the focus is on problem solving and using a probabilistic mindset.

- Course Learning Objective: adapt a useful test model used for one task to complete the task you have at hand.
- Course Learning Objective: understanding how different elements in the developmental process of a technology can effect the implementation and effectiveness of the technology.

What am I going to do? You will begin by reading the one-page prompt for this case study. In that prompt you will be given a challenge. Take time to reflect on that and make notes.

- Understand the background on how the training and development process of facial recognition technologies effect the way they work on people of different races by reading the two provided articles.
- Develop the Kaggle model to identify male and female faces. Train the model twice, once with just Wisconsin data and then once with NYC data and comparing the test accuracies.
- Determine if the Madison, WI dataset is accurate enough with the very white photodata, or if the Police Department should invest in a larger database.

Tips for success

- Comprehend what you are comparing in each train and test you conduct. Make sure you are testing on the correct model that was trained on the correct train set.

How will I know when I have Succeeded? You will meet expectations on CS-CODED BIAS Case Study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formating	<ul style="list-style-type: none"> Repository – A new github repository <ul style="list-style-type: none"> Create a new github repo for this assignment containing: <ul style="list-style-type: none"> README.md file LICENSE.md file SRC folder Slideshow Presentatation.
README.md	<ul style="list-style-type: none"> Goal: This file is what will be assessed for the assignment. You should plan to print this out and submit it in class Structure this file in such a way to be easily readable by an individual who has read the prompt Use markdown headers to divide content Include your code with any Include references
LICENSE	<ul style="list-style-type: none"> Goal: Explain to readers the terms under which they may use and share your work. The MIT license is the default recommendation
SRC folder	<ul style="list-style-type: none"> Goal: This folder contains all the source code for your project. High-level documentation for this code lives in the main level. Include any supplemental documentation as necessary, especially if it is too detailed/verbose for the overall readme.
Regression Model	<ul style="list-style-type: none"> Goal: Adapt the Kaggle logistic regression model from messy and clean to a male and female comparison and determine if the accuracy changes when the race make up of the training set changes. Model correctly outputs two accuracy percentages, one for the training mode and one for the test model. The test model is the percentage you should use for analysis. All data is provided in the Github Repository.
Slideshow	<ul style="list-style-type: none"> Goal: Create a presentation that effectively explains your analysis and clearly states your conclusions and recommendations. Clear and easy to read format. Includes statistical results. States whether the Police Department should upgrade their database and have both data and literature evidence to back up your argument. Extend on the metric of accuracy include argument on the effect of false positive and false negative results.