CS 105: Introduction to Computer Science

Prof. Thao Nguyen

Spring 2025

Materials adapted from Dave Wonnacott

Computers, Algorithms, and Society

Widely-adopted new technologies can have dramatic effects on society.

Often the harder-to-predict "secondary impacts" are dramatic.

Group activity: describe positive and negative impacts of some applications of computers.

"Predictive Policing"

Can/should we use algorithms and data in policing and the criminal justice system?

Arguments "pro" and "con" have varied, but notably include:

- "*The computer eliminates the bias that people have.*"
  - From "Predictive Policing: Using Technology to Reduce Crime", a 2013 FBI "featured article"
- *"... algorithmic techniques could reinforce existing racial biases ..."*
  - From "Predictive Policing Explained", 2020 report from the Brennan Center for Justice

One detailed example will be the focus of Lab 8.

"Machine Bias", a 2016 article from ProPublica

Lab 8 will involve algorithms to explore the use of algorithms in policing.

You'll be re-creating an analysis done by ProPublica, for their 2016 article "Machine Bias"

Elements of the lab:

- Reading the article and responding to some discussion questions
  - You'll be graded on *participation*
- Writing several Python functions (review, plus "CSV reading" library)
  - You'll be graded on *correctness* and *clarity*, as usual

This lab will be challenging in ways that differ from previous CS105 labs.

During discussions, please keep in mind that they may be more difficult for some of your classmates

Review/Introduction to relevant statistics material

When using statistics to evaluate a test, we want to look beyond "accuracy"

- *Note: positive typically means "found a problem/something of note"*
- "False Positives"
  - Test incorrectly says "positive"
  - F.P. rate = F.P.'s / total non-problem
- "False Negatives"
  - Test incorrectly says "negative"
  - F.N. rate = F.N.'s / total problem
- We'd like them both to be low
- We'd like them to be unbiased
- For details, see [this lesson from PSU]
  - Sensitivity vs. Selectivity
  - Predictive Value

		Truth: Disease (number)	Truth: Non Disease (number)	Total (number)
Test Result	Positive (number)	10 **A** (True Positive)	40 **B** (False Positive)	50 $T_{Test\ Positive}$
	Negative (number)	5 **C** (False Negative)	45 **D** (True Negative)	50 $T_{Test\ Negative}$
		15 $T_{Disease}$	85 $T_{Non\ Disease}$	100 Total

Reading a CSV file

- CSV (Comma Separated Values) is a common file format for datasets
- Python csv library

```
import csv
with open('facebook_users.csv', mode ='r') as file:
    csvFile = csv.reader(file)
    for line in csvFile:
        print(line)
```