IML Report

Kabir and Urvashi

Life Expectancy Task:

Problems:

* Encountered rows with some fields having empty entries: What to do about them?
* Encountered 0 entries: What to do about these:
  + Should make them 0.01?
  + Should leave them as is?
  + Should I just make it the mean/median/mode?

Data Preprocessing details:

* Treating “Status” (Developed or Developing countrty) as a binary 0 or 1 variable
* What to do about the year: Can I somehow bias entries that are closer to current day?
* Life Expectancy: If the main feature to be predicted is not there, I have to drop that row
* Alcohol: If NaN entrty, replace by median
* Hepatitis B: If NaN entrty, replace by average
* Measles: If NaN entry, replace by median