Designing for Thresholds

In this exercise, you will consider false positives (FP) and false negatives (FNs) and precision/recall trade-offs in the context of designing for people and society.

Imagine you are a Congolese computer scientist and you have been tasked with allocating COVID relief to all people residing in a country in East Africa. You aim to apply a ML model to complement the government's sparse data about households across the country, since phone companies have access to more information about people living in the country (see recent research to this effect). The AI system aims to ensure that COVID relief provisions are rationed equitably for each family, member, and resident of each town, with the least confusion, lost time and gas (for trucks to deliver it), and the least risk of mis-placed provisions. The ML model takes as input mobile phone data (registrations, access points, etc.) since it is assumed that almost every —though not all —households in the country own at least one mobile phone. (A household is a house where 1 or more people reside in. An example of a family residing in one household is a mother, father, one grandparent, and two young children.)

Barring for a moment the privacy concerns, ethical trade-offs, political complications, etc. of even doing this, I want you to consider the practical issues of designing and deploying the system. Even *if* the other concerns are minimal, in other words, what are the possible, *unavoidable* errors —and implications of those unavoidable errors —for an AI system for COVID relief distribution? Answer:

1. What would an example "true positive" prediction look like for this scenario? (Remember, the AI system aims to distribute relief based on mobile phone predictions of the location of households and the number of members in each household.)
2. What would a "true negative" prediction look like for this scenario? Provide an example.
3. What would a "false positive" look like for this scenario? Provide an example.

    a. Given your false positive, what are the possible social/ethical or other implications of implementation of the algorithm, if the rate of false positives is high?
4. What would a "false negative" look like for this scenario? Provide an example.
    a. Given your false negative, what are the possible social/ethical or other implications of implementation of the algorithm, if the rate of false negatives is high?
5. Given your considerations of false negatives and false positives, should the final system optimize for *precision*, or optimize for *recall?* Justify your answer in terms of *whom* (what demographics, types of people, situations, etc.) this decision would affect. (Re-watch these sections of the lecture video to remember the difference between these terms; alternatively, consult Google PAIR's article on designing around precision/recall.)