Materials and Methods

This section is separated into three parts which correspond to the three axes of the projects. Each axe is standalone and can be separated from the others, although they all interconnect in various ways.

1. Interpretability
   1. Literature review: **not really in materials and methods** - lit rev should be in background mostly (not all methods can be included in background)
      1. why it is important
      2. methods:
         1. lime; shapley ; shap; influential instances; outliers; PDP; ALE; general knowledge about what ML does well / not. Info about data used, about model used, why they match/ not, etc…**include some images and include some tables / comparisons (e.g. table: name|info|vizoutput|etc**
         2. uncertainty measures in ML: confidence intervals, aleatoric vs. (**p values etc)**
         3. visualization tools for such purposes: graphs OK but we have 100 features. Make it great. **also add a table ||||**
   2. Implementation of standalone methods to evaluate feasibility on different datasets
      1. Including performance and computational complexity / carbon footprint **this should be included in prev part 1.a.ii. in the table: add |feasibility|carbon|complexity**
   3. conduct survey with different people including clinicians (**mostly clinicians though, if we have time also different people**)
      1. **make sure prioritization of our concerns is correct according to doctors, and make this THOROUGH and experiment type procedure so its p-perfect.**
      2. Why this approach is necessary: because we need the platform to be useful to the end-users, etc…. So many ML tools but none are really helpful or tailored to the needs of clinicians, etc…
      3. Read up on what clinicians want
      4. Remind us who exactly are the end -users + they are different groups with different knowledge bases
      5. Make a questionnaire / survey + interview
      6. **report the results**
         1. **google form or else make sure you report the results !!e.g. excel/spreadsheet. for “free” answers: try to categorize to have a more insightful “overview” of the survey answers**
      7. Get some feedback, decide on what is best and timeline of adding things to platform
      8. Discuss feasibility and interest of these
   4. All of this in the context of the project + platform: we have created a platform, we want people to be able to use it
      1. How-to guide (**maybe vid too**)
      2. **the question mark thingies**
      3. Info about what models do, etc… for beginners mostly (no hardcore mathematical stuff BUT should have the resources / redirect to hardcore ish)
         1. **visualize the probabilistic space (e.g. for binary classification two groups with “divide line”)**
            1. **THIS: in general, not applied to the results! e.g. when defining a SVM just show how it divides stuff**
            2. **same for LASSO and BORUTA**
   5. Implement changes: **how much time do we have**
      1. Add textual + tabs + knowledge about how to use platform
      2. Add visualization tools in the model inspection
      3. Add some metrics in the model results page; **just add new metrics)**
      4. Add some metrics in the model choice page
   6. Changes + Error debugging in the platform
      1. Since it’s a work in progress, then obviously
      2. Also pimp out the website, change logo etc IF TIME
2. Model validation using dataset validation
   1. Lit rev **SAME AS PART 1**
   2. Why it is important, ties back to the importance of first part of project, of being able to interpret results and getting a measure of uncertainty pertaining to the results.
   3. Ebola dataset
      1. Cleaning + curation already pretty much done
      2. Make a pipeline + ML model that is comparable to the models on the platform
      3. Tune it, etc…
      4. Compare results to model results in order to establish a baseline of what is to be expected + prove that the models are “good enough” / validate their usability
   4. Conclusions on the model comparison: talk about what exactly does it prove / validate: it doesn’t validate everything obviously…
3. Platform management
   1. In coordination with Annie
   2. Oversee different projects and their future integration to the platform
   3. Organize sharing of information
      1. Platform debugging
      2. Platform improvements
      3. Communication between members
      4. Communication between members and outside organizations (IDDO, clinicians, etc…)
      5. Coordination of platform changes
   4. Evaluation of the short-term and long-term objectives of this platform and their feasibility
      1. E.g. include some other modules: interpretation (part of this thesis); imputation, clustering, sentinel/surveillance tool, multiclass labeling
      2. Include some stuff from other projects! What about new data formats (images, sounds) because some Covid detection would be great
      3. Make sure the platform stays true to its aim (although the aim might change with time) and doesn’t get bogged down with quote unquote unnecessary features (streamlined)