Meeting with annie 11\_20\_2020

* Feature imbalance: do you know if it is worth investigating, and how important is it? Can’t find much information about that.
* Talk about the feature importance plots (bi- vs. multi-variate
* Talk about the LIME importance pots
* Talk about the possibility of “making your own” test data points to see how this would influence the LIME model feature interpretation does: e.g. change a single feature value and see if it impacts the overall decision of the model and see how the feature importance is impacted overall and as a LIME model
* I talked to Ridha to get the ebola dataset
* Should try to recontact Valerie; Ludovico; rainer.
  + Before this: read up on rainer’s papers

PROGRESS:

* Done with explanation things I think. Can’t really think about more stuff to add that would really be useful. Just needs “streamlining”
  + LIME: not that great because the kernel width can change stuff and it’s not easy to “choose” an appropriate width…. So not a “definite” solution, doesn’t really help solve the problem we have !!
* However non explanation:
  + data distribution and more “data exploration” before hand
  + model choice!
  + model choice and how it separates the feature plane!
  + Should include reminders about why we are doing interpretation! Accuracy can be misleading…

Remove feature : random forest impurity since its not model agnostic

Mean might give some indication of

Bi vs multi: lines are far apart= suggests interaction !!

Ridiha: just focus on 1 dataset

Think about integration!!!!!! URGENT

Clinicolab presentation: building your own model

Doctors: have control over what they’re doing (also on the ethical side) and make conclusions about your results (and use your OWN private data that won’t be shared)

* Do check if feature importance can be negative (doubt it)
* Plot difference between bivariate and multivariate to see what it looks like (as points and lines connecting them)
* Check if signvrr1d0 is “normalized” with age and so becomes interesting only in multivariate analysis
* Data UNIQUE df to “interesting” format and beautiful
* Data distribution for non categorical variables
* SHAP better layout than LIME so checkout how to
* ML amportance of num cont value vs binary: specify that although categorical is easier to understand we would rather have continuous !
* LME: not a LINE for the mean… should be some “sum coeff” or something like that :0
* SHAP value force plot: what is the baseline, have a better explanation next time