Projektplan Mustererkennung:

Technisches Vorgehen:

* Import: (FERTIG)
  + Daten der Teilnehmer einlesen und zusammenfügen
  + Erst die Features dann die Labels
* Cleaning:
  + Fehlende Features handlen (hierzu Recherche betreiben, Paper lesen)
  + Fehlende Labels handlen (hierzu Recherche betreiben, Vaizman Paper 2 lesen)
* Preparation:
  + (Neue Features generieren, nur wenn Zeit vorhanden)
  + Labels für einzelne Sensoren bestimmen (welche Labels lassen gut mit welchem Sensor vorhersagen, z.b. Korrelation untersuchen oder simple Modelle darauf trainieren) (Hier kann man bestimmt einige Plots erstellen für die Ausarbeitung später)
  + Alternativ: Jedes Label mit jedem Modell predicten
  + Daten so aufbereiten, dass sie ins Ensamble Modelle gegeben werden können (nach Features der entsprechenden Sensoren verteilt, PhoneState in jedes Modell rein)
  + In Train-,Test- und Validationsets aufteilen (einmal so, dass Nutzerdaten gleichmäßig verteilt sind und einmal so, dass die Sets verschiedene Nutzer enthalten)
  + Features Skalieren (-m/sd)
* Modelling:
  + Modelle für einzelne Komponenten des Ensambles bestimmen (xgboost, NN, logistic Regression, kNN, RandomForest, Catboost) (oder für jeden Sensor das selbe Modell?) (Hier können wieder Plots für die Ausarbeitung erstellt werden)
  + Ensamble Modell bauen mit LFL Late Fusion using learned weights (es sei denn die einzelnen Komponenten predicten verschiedene Sachen) (pro Label ein Modell) (hier Recherche betreiben, wie schummelt man nicht beim letzten Modell?) (Auch hier wieder Plots bezüglich Loss etc. generieren)
* Evaluation:
  + Balanced Accuracy für alle Labels bestimmen (Tabelle bauen für Ausarbeitung)
  + Ergebnisse mit denen von Vaizmann vergleichen (Hier kann man bestimmt auch irgendeine Grafik/Tabelle für die Ausarbeitung gewinnen)
  + Evaluieren, welche Labels gut zu predicten waren und warum
* Bonusaufgaben:
  + XAI (Skater oder Shap über Ensamble Modell ziehen und ein paar Plots ausgeben lassen)
  + Nutzer predicten (einfach XGBoost über die entsprechenden Daten ziehen)

Offene Fragen:

* Wie gibt Sören die Daten ein bzw. wie müssen wir unser Ergebnis strukturieren, damit er es benutzen kann?
* Pip freeze / virtual env
* Kann man die Bonusaufgabe XAI in der normalen Aufgabe integrieren?
* Welche Schriftgröße soll man benutzen oder sollte sie einfach einheitlich sein?
* Was machen wenn Rechenleistung nicht reicht? Paramatertuning weglassen?