Advanced clustering/Module 5 proposal

For the advanced clustering section of this project, we intend to pursue KMeans++ along with some other changes to our ODM & OCRM to achieve an optimal result. The primary area for improving our ODM right now is combating overfitting. One idea we have for this is to train ensemble methods, such that there are greater than 3 different models detecting objects in the ERD, and then we look at the combined result of all of them. We would use the combined result of the multiple methods by using the labels that appear the most, the average pixel locations for things such as box / circle identification, average scores, and any other features within the models that we want to optimize. It is important to clarify that in doing this, we will only be using an odd number of methods, such that it's not possible to have an even split in terms of what label appears the most. As far as improving the model, we also want to do non max suppression, where in the case that a model places a box on whitespace, we tell it that there is no overlap between that whitespace and a known ground truth value. For the actual KMeans++, we intend to vary what method of distance evaluation we use and compare results. We intend to vary the distance measurement strategy by using different types of norms, such as Manhattan distance (L1 norm), L3, L4 norm, etc.