**Unsupervised**

**Dimension Reduction**

A PCA on all the data: OD, d/dx(OD), ln(d/dx(OD)), od alignment (unsure if time warping is possible for PCA).

A PCA on the highest glucose concentration: OD, d/dx(OD), ln(d/dx(OD)) od alignment (unsure if time warping is possible for PCA).

**Clustering**

A clustering algorithm\* all the different glucose concentrations (to show the glucose concentrations separate): OD, d/dx(OD), ln(d/dx(OD)) with time warping, vs od alignment.

A clustering algorithm\* on the highest glucose concentration to show that strains can be separated, albeit by a latent variable I am not sure of: OD, d/dx(OD), ln(d/dx(OD)) with time warping, vs od alignment.

\*Kmeans, Kmedoids, potentially more advanced techniques; time permitting.

**Supervised**

**All Data**

A shallow learning model x 2 (SVM like Zhang, and Forest?) on all the data to see if glucose can be predicted: OD, d/dx(OD), ln(d/dx(OD)) with time warping, vs od alignment.

A deep learning model to see if glucose can be predicted: OD, d/dx(OD), ln(d/dx(OD)) with time warping, vs od alignment.

**Highest Glucose**

A shallow learning model x 2 (SVM like Zhang, and Forest?) on the highest glucose to see if strain can be predicted: OD, d/dx(OD), ln(d/dx(OD) with time warping, vs od alignment.

A deep learning model to see if strain can be predicted: OD, d/dx(OD), ln(d/dx(OD)) with time warping, vs od alignment.

**\*Potential\***

Using the PCA metrics as features in both the clustering and supervised.

***Datasets I need:***

|  |  |
| --- | --- |
| **Glucose 0.01** | **All** |
| OD, | OD, |
| d/dx(OD) | d/dx(OD) |
| ln(d/dx(OD)) | ln(d/dx(OD)) |