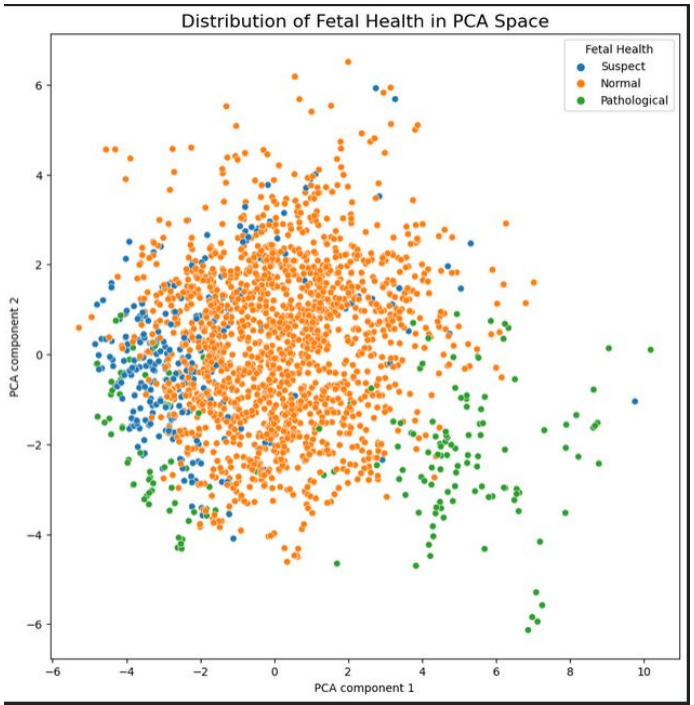
Fetal-health-classification(Directly copied from kaggle)

Tony Slu -> (Jupyter notebook)

Visualization of Data Labels per PCA component explainability



- Lay over of label classes over the 2 principal component vectors to confirm explainability of components
- We can see that there isn't a particular vector that diverges from the major classes in this dataset and can assume that the classes contribution to the over distribution of the dataset is fairly well distributed

Model Description

- Random Forest based model
  - Ensemble of decision trees
  - Negative log likelihood regressor
  - Gradient Descent based
  - First egress expansion
- XGBoost classifier
  - Objective function reg:squarederror
  - Second degree taylor series expansion
  - hessian/gradient
- CatBoost classifier
  - CatBoost, uses the concept of ordered boosting, a permutation-driven approach to train model on a subset of data while calculating residuals on another subset, thus preventing target leakage and overfitting

Model Performance

Model	Accuracy	AUC	F1-Score(Macro)
Random Forest	0.93	0.98	0.9
Xgboost	0.94	0.98	0.9
CatBoost	0.96	0.98	0.94

Confusion Matrix

- From this confusion matrix, we can also see a very good model harmonization between precision and recall of classes