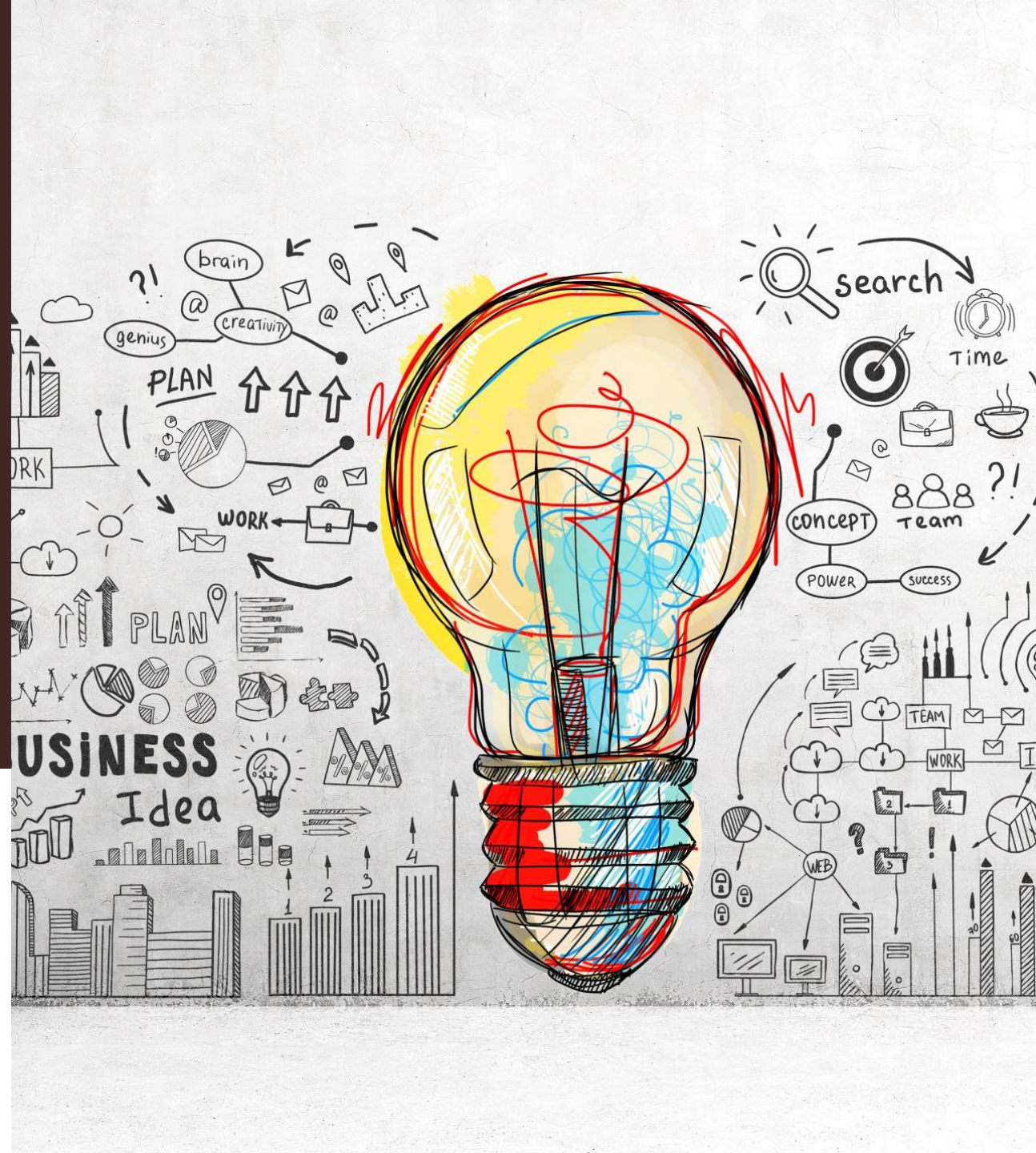# Applied Machine Learning Spring 2024 Group 9
Home Credit Default Risk

Paul Miller, Glen Colletti, Alex Bordanca

Project Description

- Home Credit Default risk
  - Predict from a variety of non-traditional credit metrics if a loan will be repaid
  - Kaggle competition for Home Credit company
    - Provide loans to those without traditional financial history
    - Mitigate risk
  - Preprocessing
  - Feature selection
  - Feature Engineering
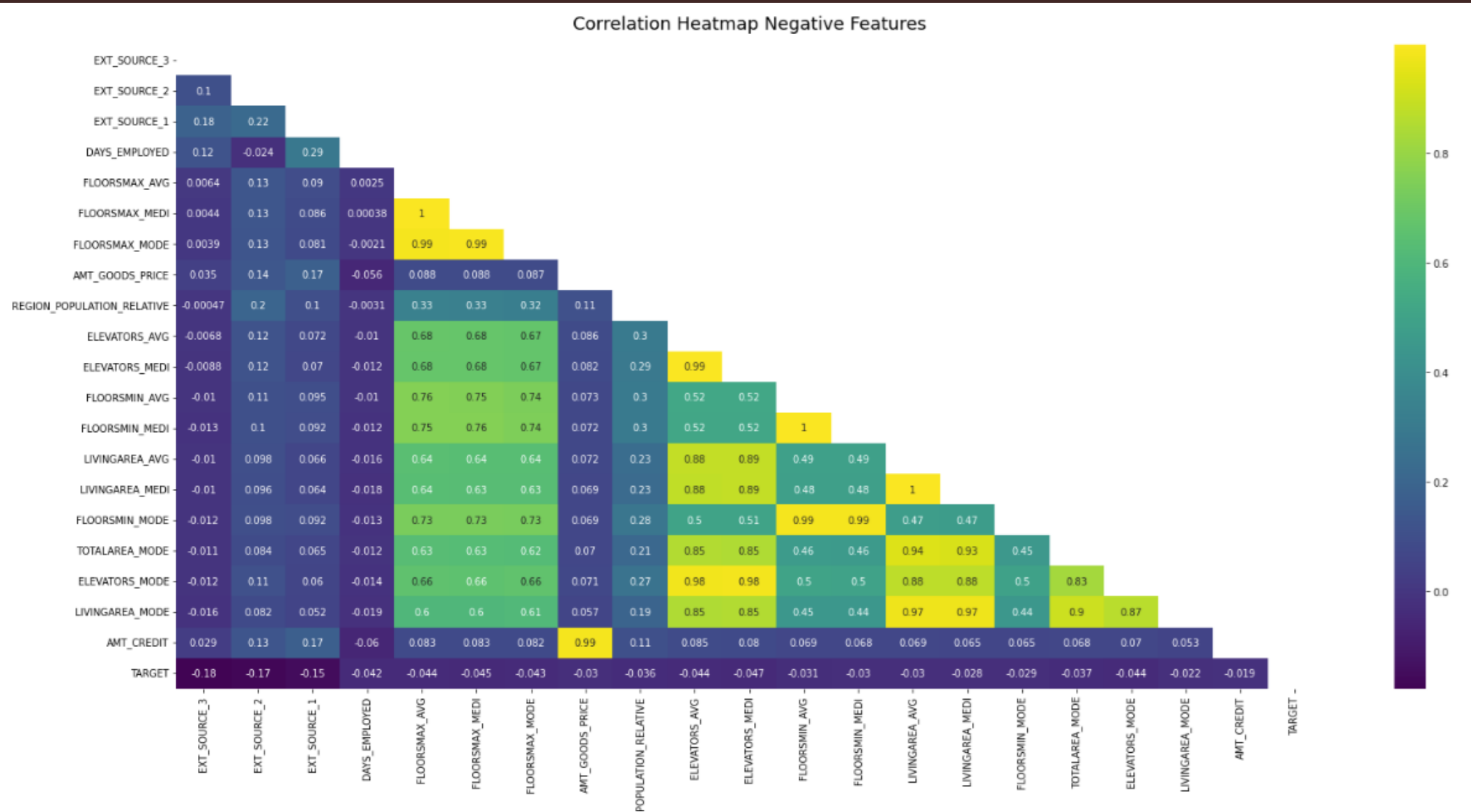 - Implement XG Boost, Logistic Regression, KNN, Neural Network classifier models

Imbalanced target oversampling

In original data: 0 class = 92%, 1 class = 8%
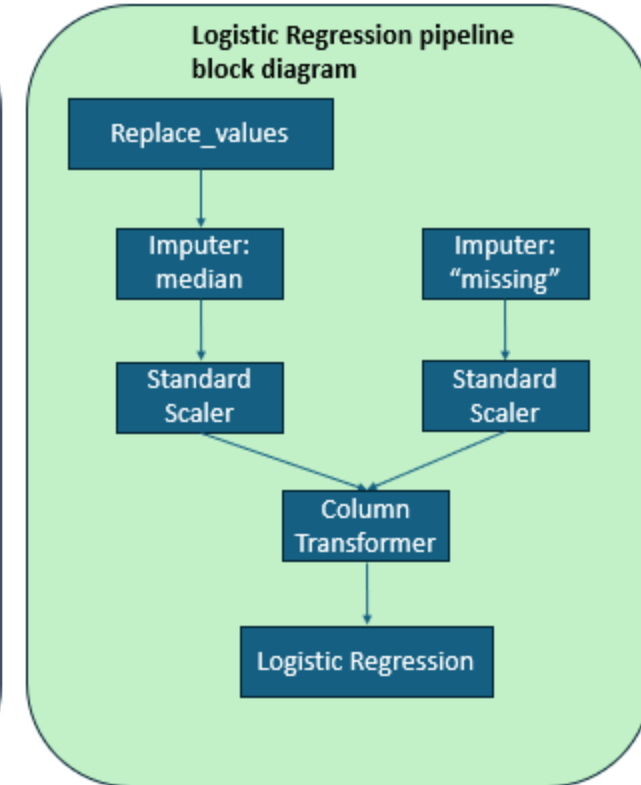Oversampling minority class to create balanced distribution
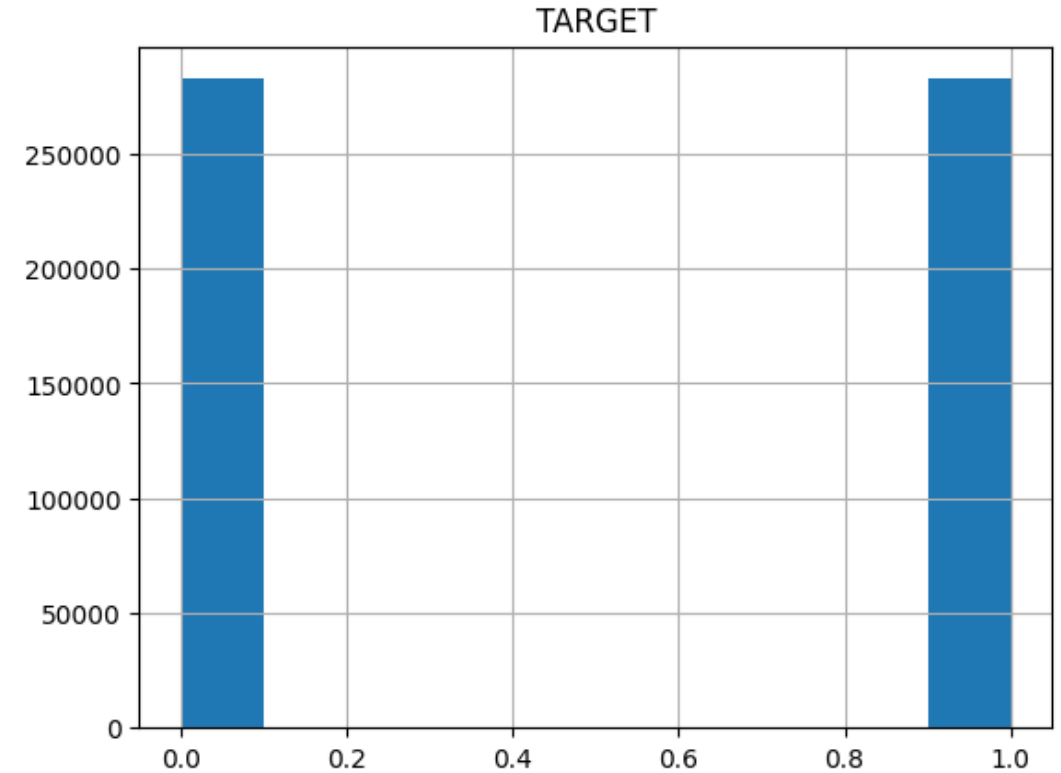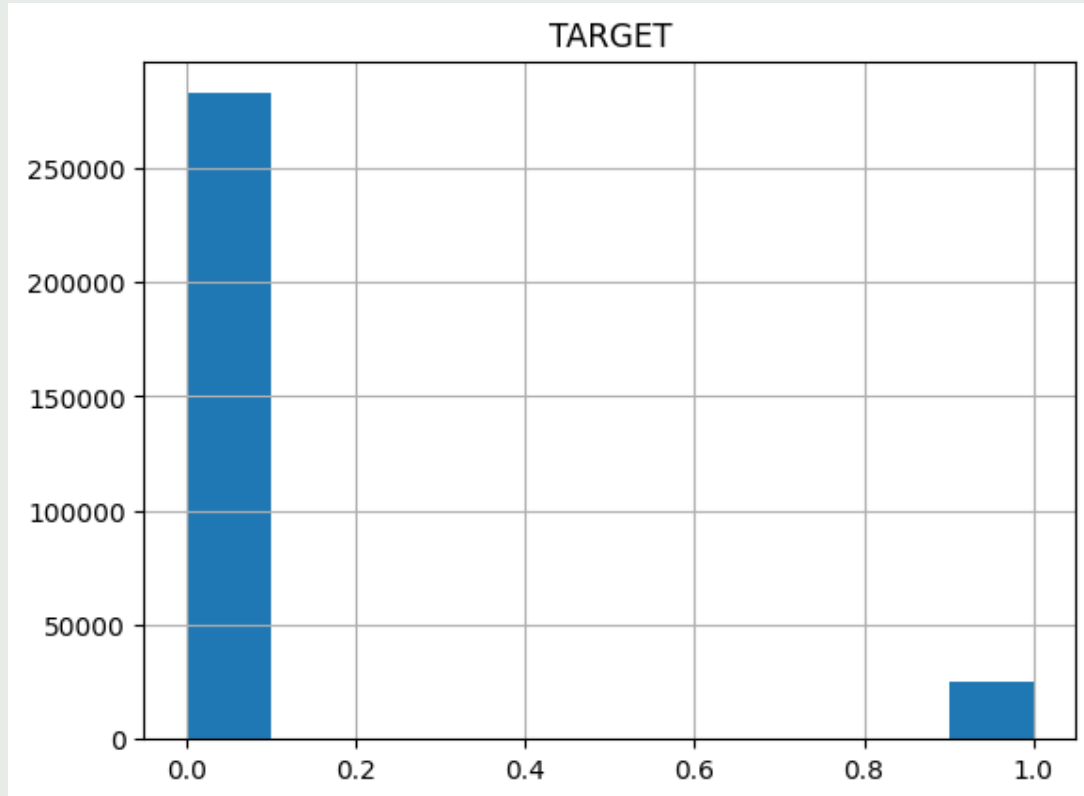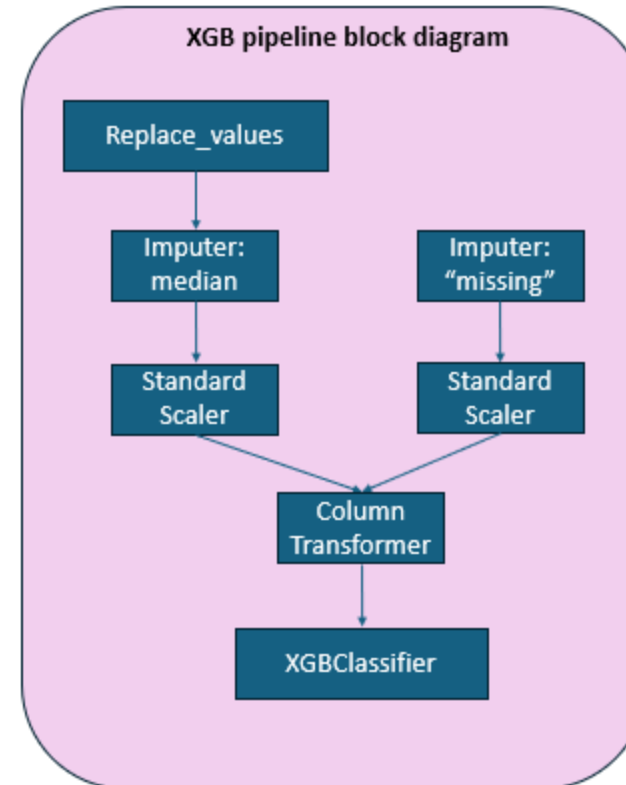
Multicollinearity between features (data leakage)

Correlation Heatmap Negative Features
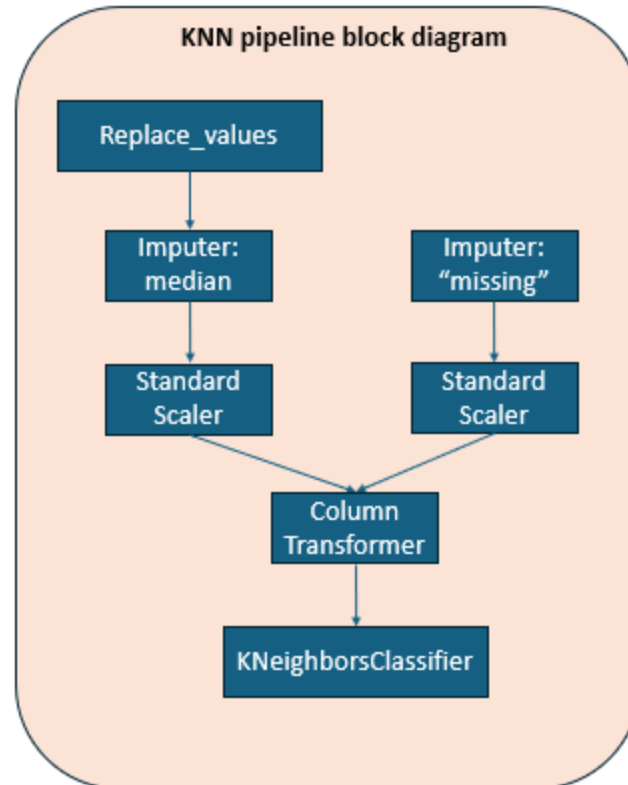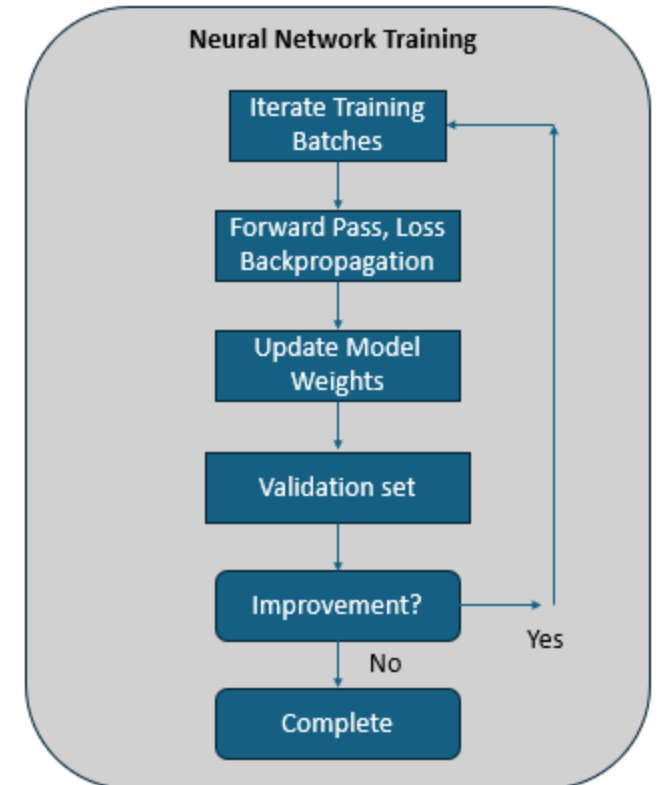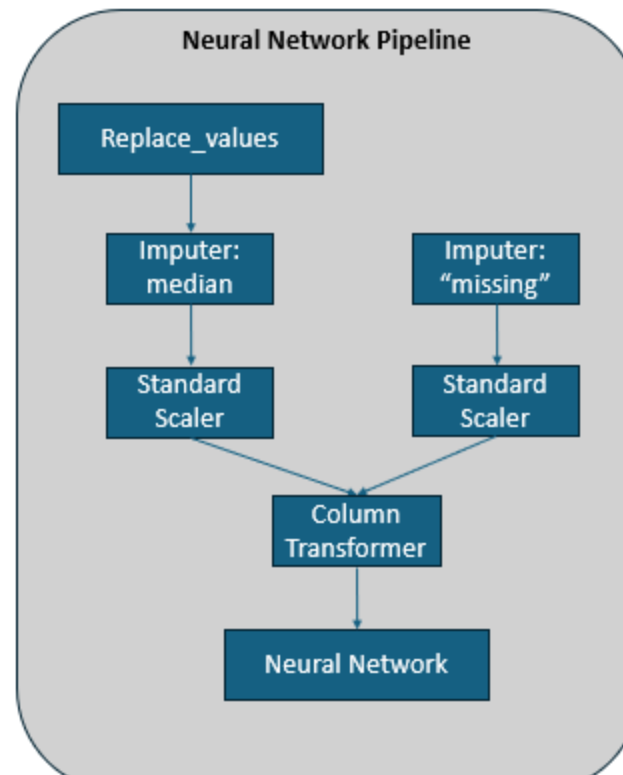
Modeling Pipelines

- Pipeline Updates
  - Add feature transformer for spurious values
  - Added grid search for hyper parameter tuning

Modeling Pipelines continued

- Neural Network
  - Sigmoid and Rectified Linear Unit activation function
- 34 node initial layer
- 2 node final layer
- Variable hidden layers
- 92% ceiling without oversampling
- 60% ceiling with oversampling

Results

- F1 scores and imbalanced target data
- Neural Network had highest ROC AUC scores

	exp_name	Train ROC AUC	Valid ROC AUC	Test ROC AUC	Train F1	Valid F1	Test F1
0	XGBoost_baseline	0.520	0.508	0.500	0.079	0.032	0.000
1	XGBoost_tuned	0.500	0.500	0.500	0.000	0.000	0.000
2	LogReg_baseline	0.637	0.633	0.655	0.000	0.000	0.000
3	LogReg_tuned	0.637	0.633	0.655	0.000	0.000	0.000
4	KNN_baseline	0.534	0.500	0.505	0.129	0.015	0.031
5	KNN_tuned	0.534	0.500	0.505	0.129	0.015	0.031

	Architecture string	Optimizer	Epochs	Train accuracy	Valid accuracy	Test accuracy	test F1 score	test ROC_AUC score
0	34-50-50-25-2	<class 'torch.optim.adadelta.Adadelta'>	5	56.599999999999994%	57.4%	57.4%	57.099999999999994%	57.4%
1	34-50-50-25-2	<class 'torch.optim.adadelta.Adadelta'>	5	56.6%	57.4%	57.4%	57.1%	57.4%
2	34-100-100-2	<class 'torch.optim.adadelta.Adadelta'>	10	59.5%	59.7%	59.5%	59.5%	59.5%
3	34-100-100-2	<class 'torch.optim.sgd.SGD'>	10	59.9%	60.2%	59.9%	59.9%	59.9%
4	34-34-34-17-2	<class 'torch.optim.sgd.SGD'>	10	59.4%	59.7%	59.5%	59.5%	59.5%
5	34-200-100-2	<class 'torch.optim.sgd.SGD'>	10	60.1%	60.4%	60.2%	60.2%	60.2%
6	34-200-100-50-2	<class 'torch.optim.sgd.SGD'>	10	60.0%	60.1%	59.9%	59.9%	60.0%
7	34-200-100-50-2	<class 'torch.optim.sgd.SGD'>	10	60.0%	60.1%	59.9%	59.9%	60.0%
8	34-200-100-50-2	<class 'torch.optim.sgd.SGD'>	10	60.0%	60.1%	59.9%	59.9%	60.0%