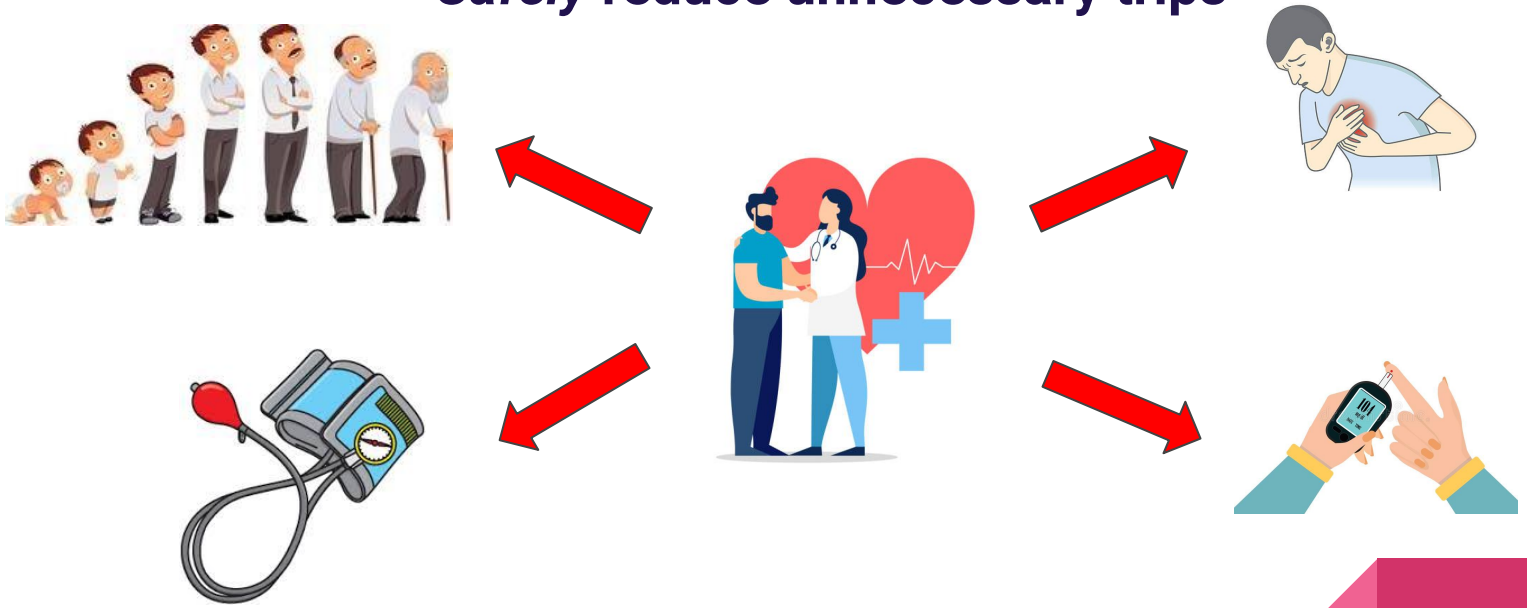
ML1000 - Project 1
Heart Disease

Health Care for remote and Northern communities often means Travel

- Difficult, many locations by air only most of the year
- Expensive
- Separates patients from their communities and families
- Sometimes unnecessary

Create a diagnostic tool using Machine Learning

Safely reduce unnecessary trips



First Target: Heart Disease

Heart Disease Dataset for Model Training

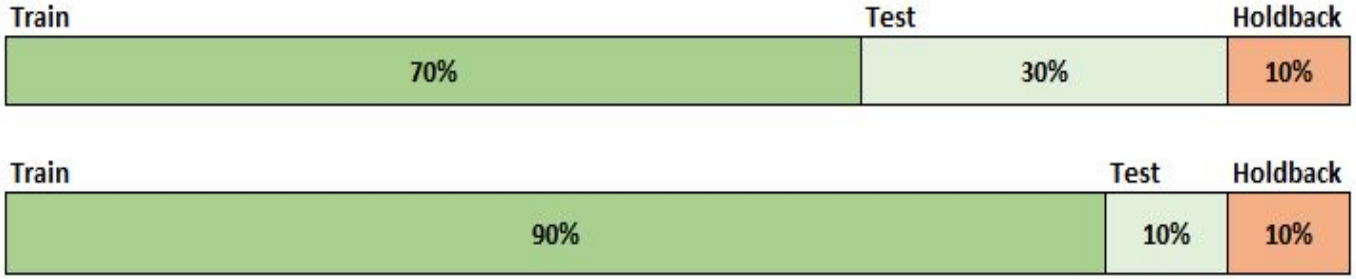
297 Records

14 Features, incl. Balanced Target (46% Positive)

Age	Gender	Chest_Pain	Blood_Pressure	Cholesterol	Blood_Sugar	Rest_ECG	Max_Heart_Rate	Exercise_Angina	ST_Depression
69	Male	Typical	160	234	120mg+	LV Hypertrophy	131	No	None
69	Female	Typical	140	239	<120mg	Normal	151	No	High
66	Female	Typical	150	226	<120mg	Normal	114	No	High
65	Male	Typical	138	282	120mg+	LV Hypertrophy	174	No	Low
64	Male	Typical	110	211	<120mg	LV Hypertrophy	144	Yes	High
64	Male	Typical	170	227	<120mg	LV Hypertrophy	155	No	Low
63	Male	Typical	145	233	120mg+	LV Hypertrophy	150	No	High
61	Male	Typical	134	234	<120mg	Normal	145	No	High

ST_Depression highly skewed numeric, one third zero
Converted to Categorical feature, None, Low, and High

Data Setup



- ➔ Train/Test Split
- ➔ Numeric binning
- ➔ Principle Component Analysis
- ➔ Outliers
- ➔ Multicollinearity
- ➔ Normalization

Model Creation - Folds

Model Creation - Optimization

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

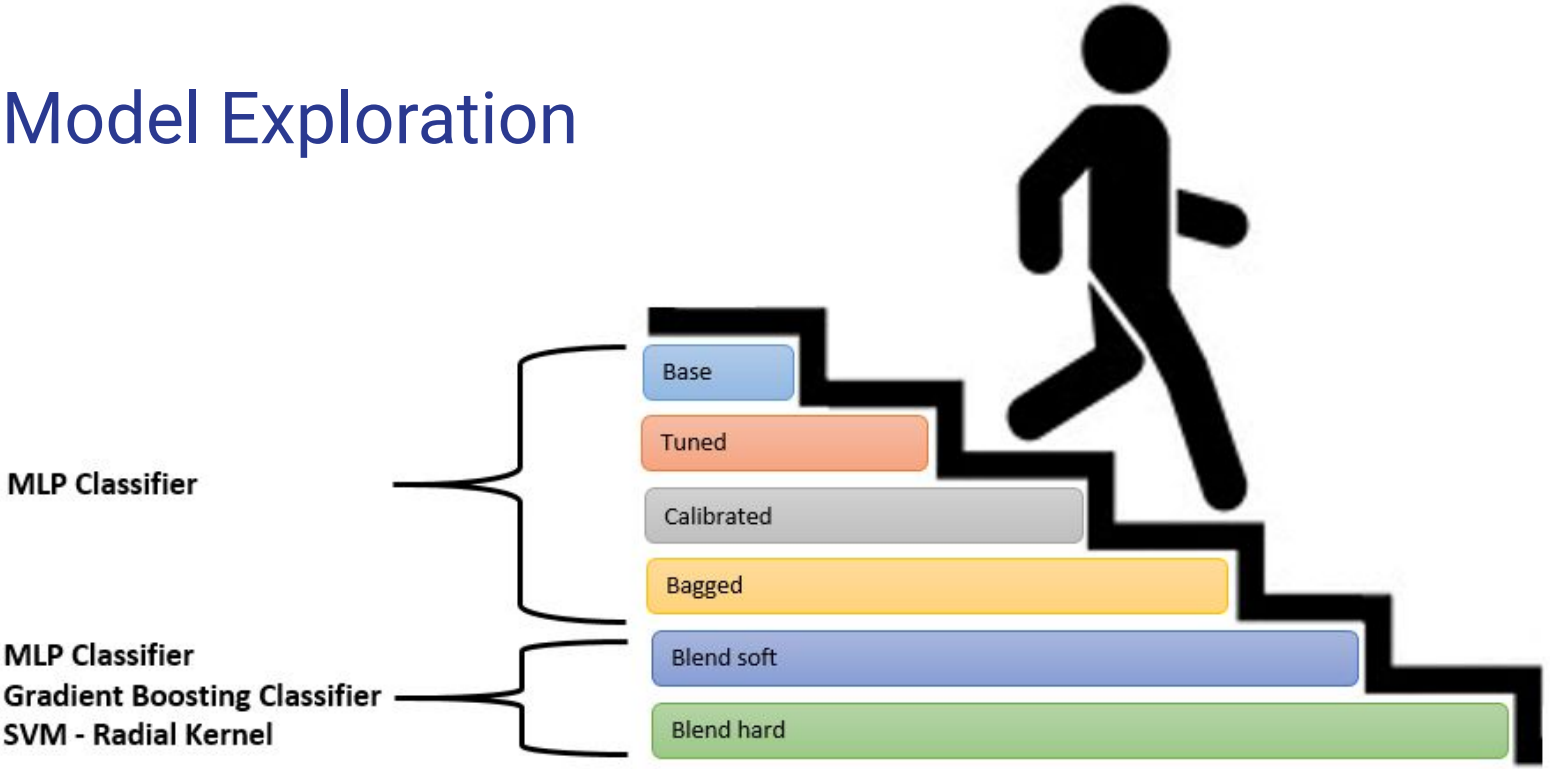
True negative Predicted negative Actual negative	**False positive** Predicted positive Actual negative
False negative Predicted negative Actual positive	**True positive** Predicted positive Actual positive

Model Analysis/Process

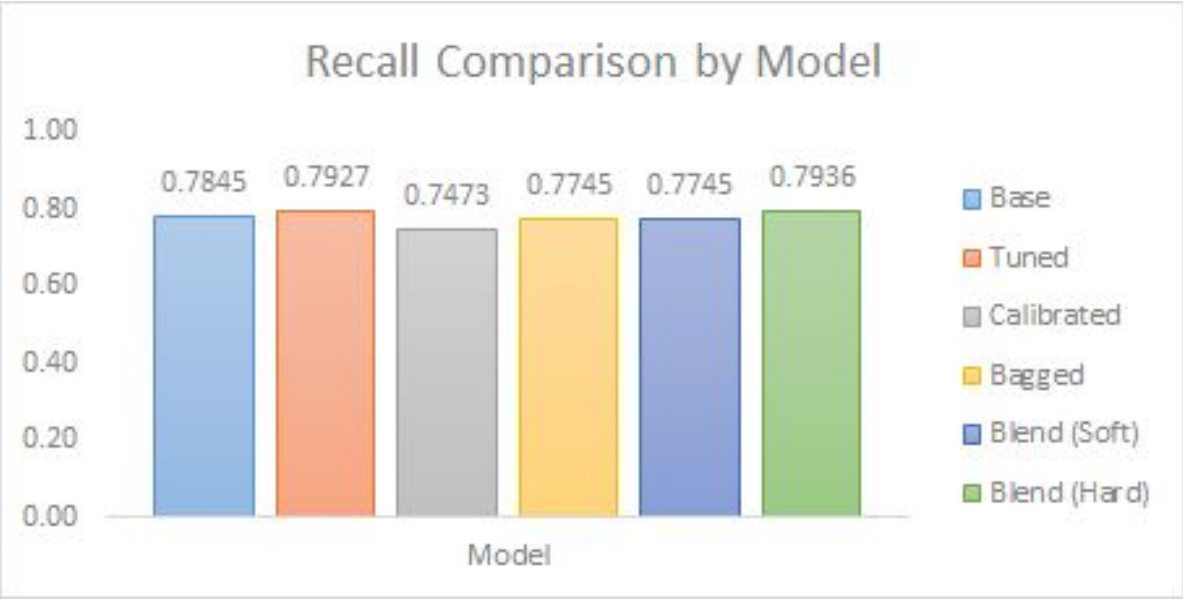
	Model
mlp	MLP Classifier
gbc	Gradient Boosting Classifier
rbfsvm	SVM - Radial Kernel
lightgbm	Light Gradient Boosting Machine
lr	Logistic Regression
et	Extra Trees Classifier
qda	Quadratic Discriminant Analysis
knn	K Neighbors Classifier
ridge	Ridge Classifier
lda	Linear Discriminant Analysis
dt	Decision Tree Classifier
rf	Random Forest Classifier
svm	SVM - Linear Kernel
nb	Naive Bayes
gpc	Gaussian Process Classifier
ada	Ada Boost Classifier

	Model	Accuracy	AUC	Recall	Prec.
mlp	MLP Classifier	0.7976	0.9052	0.7845	0.8042
gbc	Gradient Boosting Classifier	0.8024	0.8819	0.7845	0.8079
rbfsvm	SVM - Radial Kernel	0.8152	0.8979	0.7836	0.8261
lightgbm	Light Gradient Boosting Machine	0.8148	0.8828	0.7818	0.8219
lr	Logistic Regression	0.8067	0.9007	0.7655	0.8226

Model Exploration

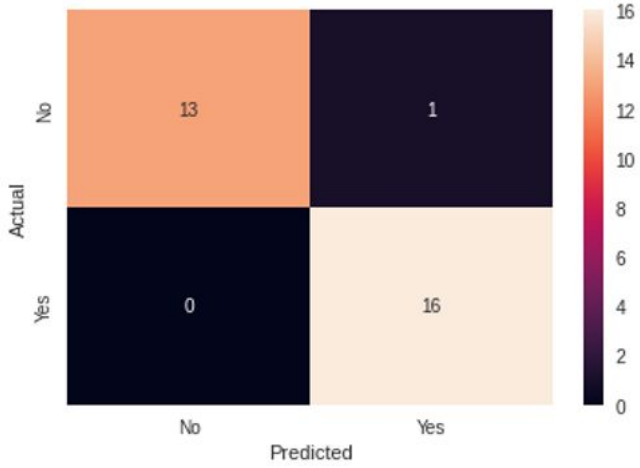


Model Comparison



Chosen Model

```
MLPClassifier(activation='identity', alpha=0.5, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=[100], learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=500,
              momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
              power_t=0.5, random_state=123, shuffle=True, solver='adam',
              tol=0.0001, validation_fraction=0.1, verbose=False,
              warm_start=False)
```



Accuracy: 0.9667

Recall: 1.0000

Precision 0.9412

Conclusion

Thank You