Heart Disease Classification Model

by M. Boyd-Vasiliou, A. Weber, Z. Zaidi

Background

Delivering health care services in Nunavut is challenging for several reasons: size of the territory, dispersion of the small population, weather, and reliance on air transportation. The territory's 25 communities are isolated and spread across the largest jurisdiction in Canada.

Nunavut's communities are accessible year-round only by air. While commercial flights are used for non-urgent travel, air ambulances are required for medical emergencies. Depending on the services available in their local community, patients are often sent to larger centres for treatment of conditions such as stroke, heart attack, or cancer. Within the territory those centres are Iqaluit, Rankin Inlet and Cambridge Bay, but specialized care can usually be obtained only by travel to Yellowknife, Ottawa, Winnipeg, or Edmonton. In 2016 transportation accounted for almost 17% of the total health expenditure in the territory.

Outside of the three centres with physician-staffed facilities noted above, primary health care is locally administered by a Community Health Nurse, with remote physician consultation. It is this team that is responsible for preliminary diagnosis, and making a decision for transport if needed. Transport is a huge disruption - beyond the obvious financial cost it separates the patient from their community and family, and often none are able to accompany the patient and provide support.

It is hoped that with better diagnostics the health care team could make transport decisions with greater confidence and minimize unnecessary trips - a win-win for both the patient and the capacity of the health system. The territory believes data science can provide some of these tools, and they've engaged our team to develop a proof-of-concept on heart disease to evaluate the feasibility of this approach.

Proof-of-concept

Our team will develop a diagnostic classification model for heart disease, and this will be demonstrated with a web application. The functionality of the system and performance of the model will be demonstrated in a conference-room verification for an evaluation team including nurses, doctors, health administrators and the project sponsor.

Key Metrics and Functionality

Performance

- Model must have a false-negative rate <5%, ie. a healthy diagnosis of a sick individual
- Model must have a false-positive rate <30% while achieving the goal above, in order to identify a significant number of healthy individuals who can safely avoid transport, which is the primary objective.

Reliability

- Model should have documentation that establishes confidence that its performance is robust and repeatable.

Data Capture

- System must log its inputs and predictions for performance analysis and future training data

Usability

- System interface needs to demonstrate capability only, proper User Acceptance testing would follow in a pilot project if the proof-of-concept is satisfactory

Data Overview

This dataset is the Cleveland variant of the UCI Heart dataset, it was downloaded from here: https://www.kaggle.com/cherngs/heart-disease-cleveland-uci. The dataset consists of 297 observations with 14 columns of data.

The dataset columns were renamed to more descriptive and easily understandable names. Column renaming does not impact any of the model analysis but provides an easier understanding and usefulness in analysis, graphing, and review.

Data Dictionary

#	Original Col. Name	New Col. Name	Description	Notes/Details
1	age	Age	Age in years	Numeric
2	sex	Gender	Client gender	Categorical data: ➢ Male ➢ Female
3	cp	Chest_Pain	Chest pain type	Categorical data: ➢ Typical ➢ Atypical ➢ Not Anginal ➢ Asymptomatic
4	trestbps	Blood_Pressure	Resting blood pressure (in mm Hg on admission to the hospital)	Numeric
5	chol	Cholesterol	serum cholesterol in mg/dl	Numeric
6	fbs	Blood_Sugar	fasting blood sugar	Categorical data: ➢ <120mg ➢ 120mg+
7	restecg	Rest_ECG	resting electrocardiographic results	Categorical data: ➢ Normal ➢ ST Abnormality ➢ LV Abnormality
8	thalach	Max_Heart_Rate	maximum heart rate achieved	Numeric
9	exang	Exercise_Angina	exercise induced angina	Categorical data: ➢ No ➢ Yes

10	oldpeak	ST_Depression	ST depression induced by exercise relative to rest	Categorical data: ➢ None ➢ Low ➢ High
11	slope	ST_Slope	the slope of the peak exercise ST segment	Categorical data: ➢ Flat ➢ Upslope ➢ Downslope
12	ca	Marked_Vessels	number of major vessels (0-3) colored by flourosopy	Categorical data: ➢ 0 ➢ 1 ➢ 2 ➢ 3
13	thal	Thallium	Results of Thallium stress test	Categorical data: ➢ Normal ➢ Fixed Defect ➢ Reversible Defect
14	condition	Heart_Disease	Target Variable Signifies whether client has heart disease or not.	Target variable Categorical data: ➢ No ➢ Yes

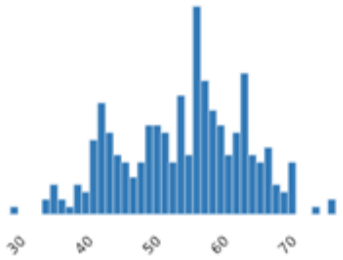
Data Exploration

Dataset statistics		Variable types	
Number of variables	14	**Numeric**	5
Number of observations	297	**Categorical**	9

Numeric Data

Age

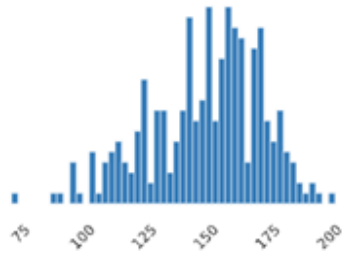
Minimum	29
Maximum	77
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%



Age has a relatively normal distribution, with a mean of 55 years. Older than an average population, which makes sense as this group have all been examined in hospital for possible heart conditions. It is interesting to note the double-hump, with a secondary density of patients in their early forties.

Max_Heart_Rate

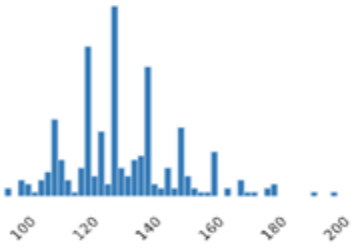
Minimum	71
Maximum	202
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%



A normal distribution, with a mean of 150. The min value of 71 looks like an outlier, this is supposed to be an exercise test and that figure suggests exercise was not performed, whatever the reason.

Blood_Pressure

Minimum	94
Maximum	200
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%

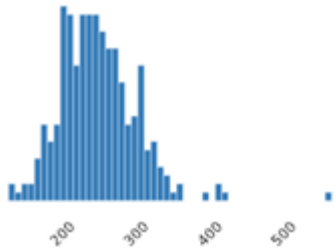


Normal distribution, mean of 132. Also the high level markers indicate some patients under considerable duress.

Cholesterol

Normal distribution, mean of 247. The max number of 564 looks off the scale, but figures above 500 are within the range of expected results, it's simply considered a very high score

Minimum	126
Maximum	564
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%



ST_Depression

Lopsided distribution, and a third of the values are zero, which is a normal, healthy condition. The ST_Depression is measured by exercise relative to rest, and the test is supposed to stop when the gap reaches 2mm. There are values in the data that go well beyond, up to 6.2mm. If we hard-cap the max value at 2, we'd in effect create a bin at 2, similar to the naturally occurring bin on the other side of the scale. What we did was create three categories for ST_Depression, as follows, and it is analyzed further with the other categorical features.

Minimum	0
Maximum	6.2
Zeros	96
Zeros (%)	32.3%
Negative	0
Negative (%)	0.0%

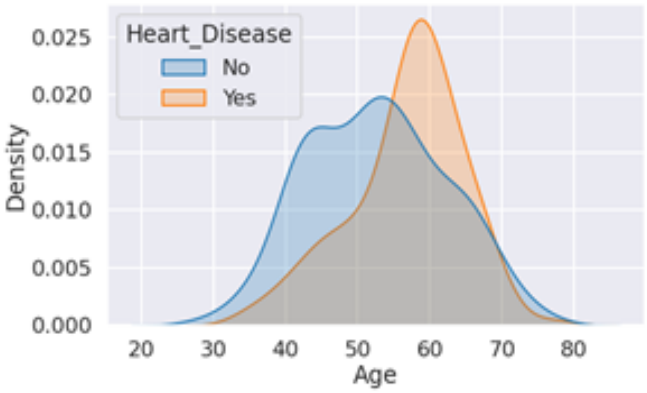
Category	ST Depression Range	Qty
None	0	102
Low	0.1 - 1.5	121
High	> 1.5	74

Correlation of Numeric features to Heart Disease

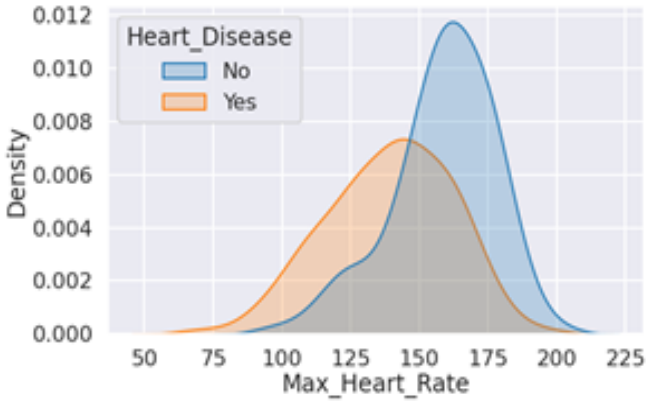
Mean Value of Numeric features compared across presence of Heart Disease				
Heart_Disease	Age	Blood_Pressure	Cholesterol	Max_Heart_Rate
No	53	129	243	159
Yes	57	135	252	139
All	55	132	247	150

This chart above shows the mean scores of patients grouped by whether they had heart disease. As expected, age, blood pressure and cholesterol all averaged higher in the disease group. Conversely, a higher value for Max Heart Rate indicates a healthier heart (able to beat faster), and so its correlation goes in the opposite direction, with a lower score associated with the disease group. It is also the strongest correlation of the four, as can be seen in the following graphs.

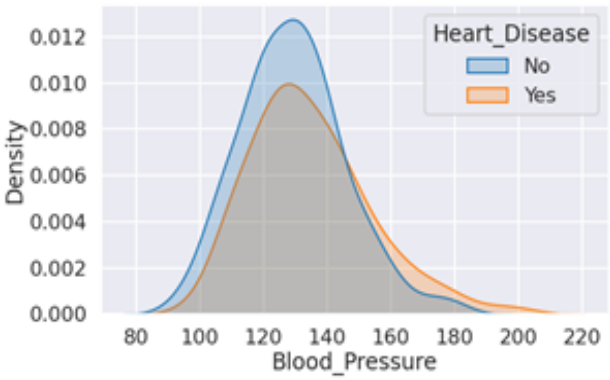
Age vs Disease



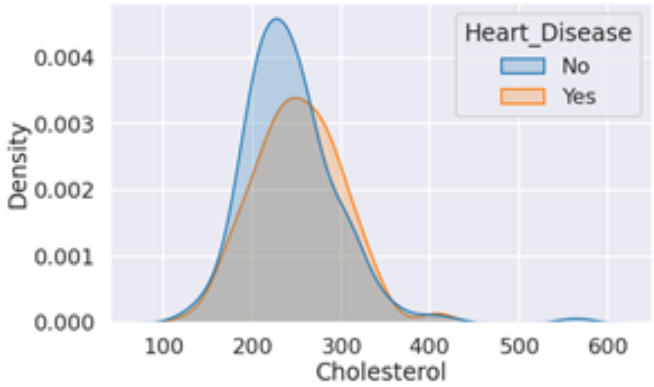
Max_Heart_Rate vs Disease



Blood_Pressure vs Disease



Cholesterol vs Disease

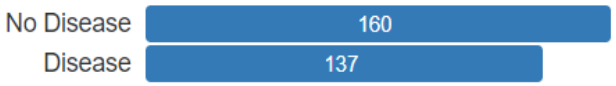


And as expected, Age and Max Heart Rate showed the most correlation with each other, with younger patients having the higher Max Heart Rate, so there's a general downslope to the scatterplot, with a trend away from Heart Disease:

Max _Heart_Rate vs Age, Disease

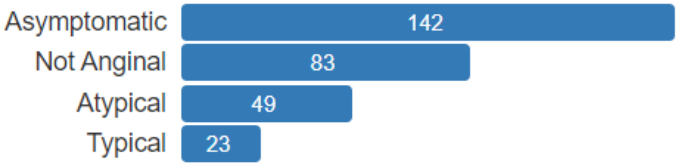
Categorical Data

Heart_Disease



46% of the patients in the dataset were identified as having a heart disease.

Chest_Pain



There are three classes of angina (chest pain) according to these criteria:

- Chest pain occurs around the substernal portion of the body
- Pain is experienced after induction of emotional/physical stress
- The pain goes away after taking nitroglycerine and/or a rest

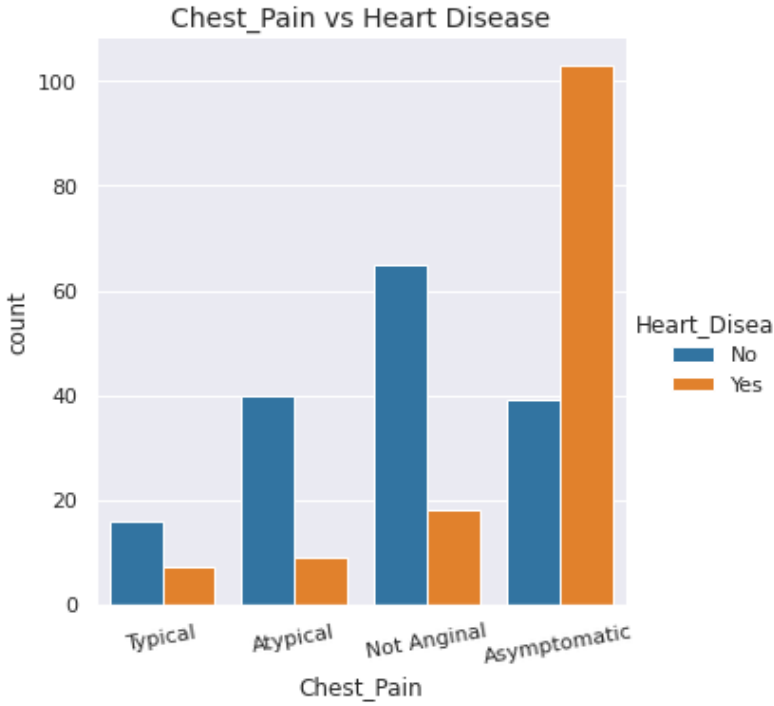
Typical: All criteria present

Atypical: Two of three criteria

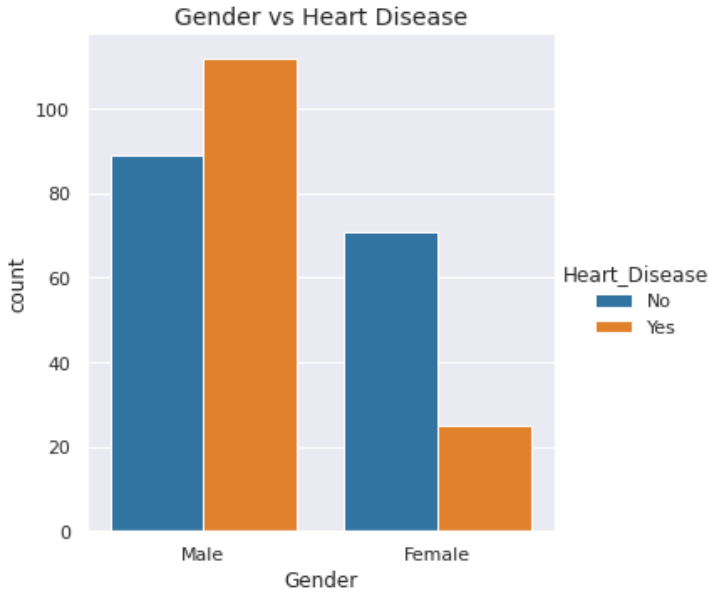
Non-Anginal: One criteria satisfied

Asymptomatic: None of the criteria are satisfied

Our data has a very strong correlation between Chest_Pain and Heart Disease, with the category Asymptomatic representing almost half of the records

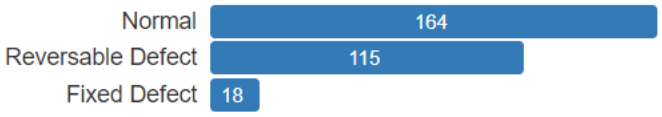


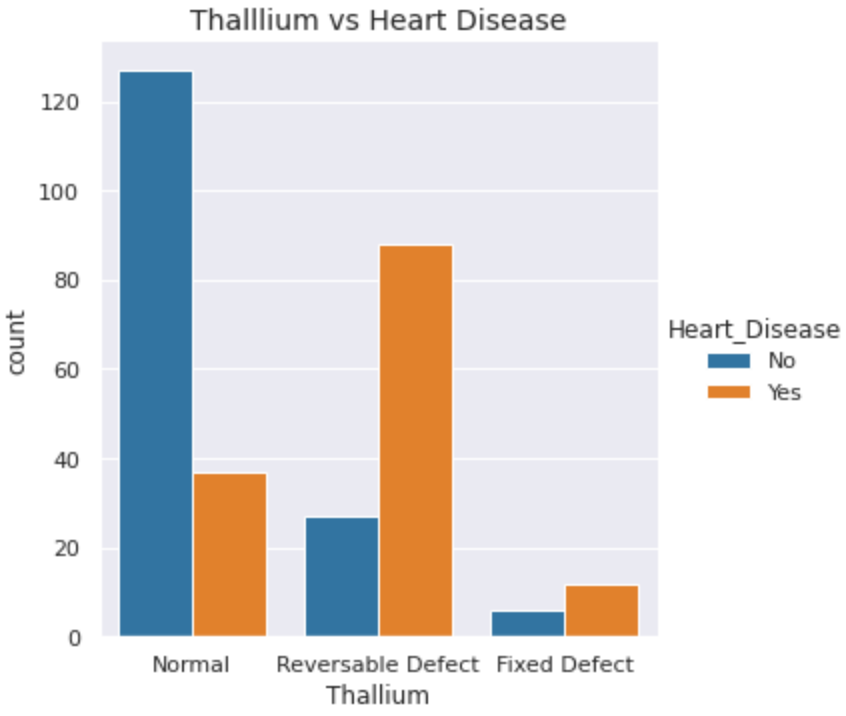
Gender



Men outnumber women in the sample cohort by a 2-to-1 margin, remember this is not a random sample, but patients admitted to hospital to be checked for heart disease. Even within this group however, there is a large discrepancy in the probability of having the disease, 55% for the men, and only 26% for the women.

Thallium





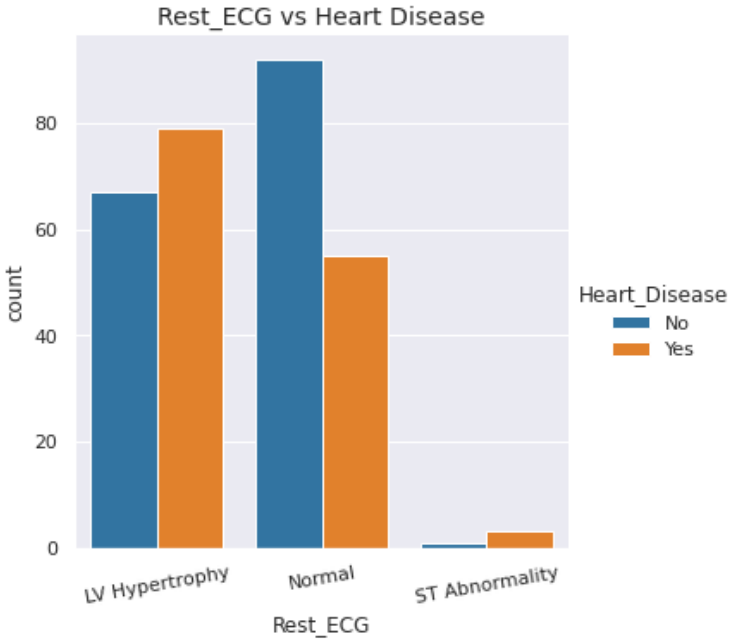
A thallium stress test is a nuclear imaging test that shows how well blood flows into your heart while you're exercising or at rest.

Normal: heart tissue is able to absorb thallium

Reversible Defect: heart tissue is unable to absorb thallium only under the exercise portion of the test

Fixed Defect: heart tissue can't absorb thallium both under stress and in rest

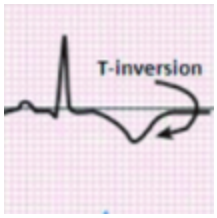
Rest_ECG



Shape of Resting electrocardiogram wave:

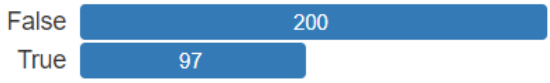
Normal: normal

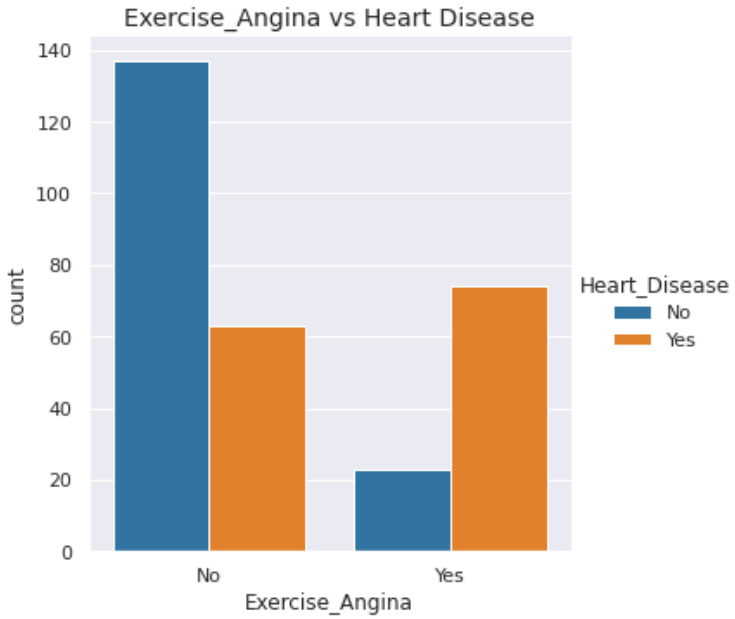
ST Abnormality: having ST-T wave abnormality such as T-wave inversions



LV Hypertrophy: showing possible left ventricular hypertrophy

Exercise_Angina

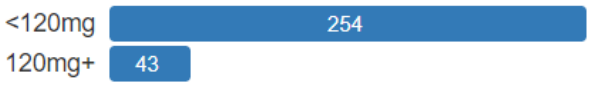


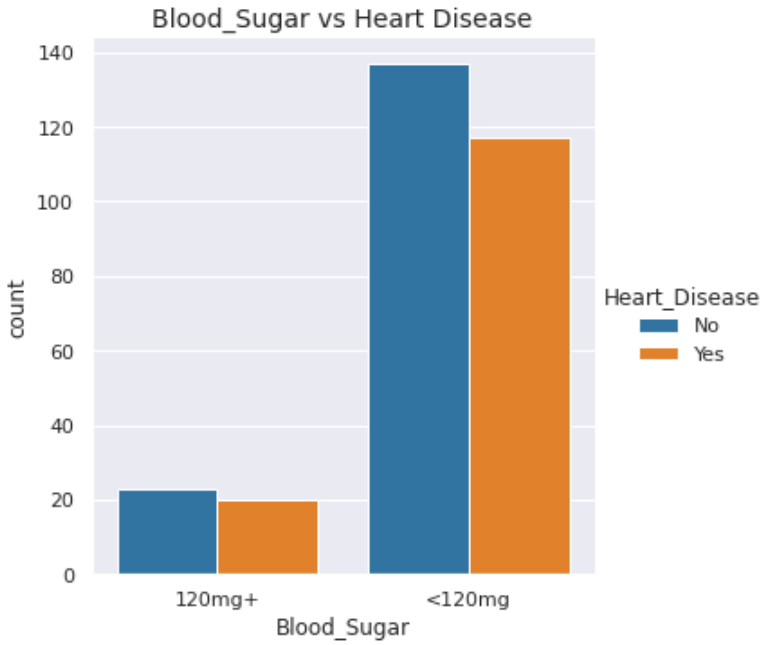


Patient experiences chest pain after exercise

Correlation is very strong, 76% of patients with the symptom had heart disease, compared to only 32% of patients without

Blood_Sugar

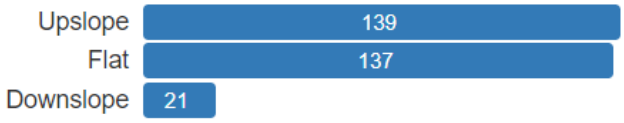




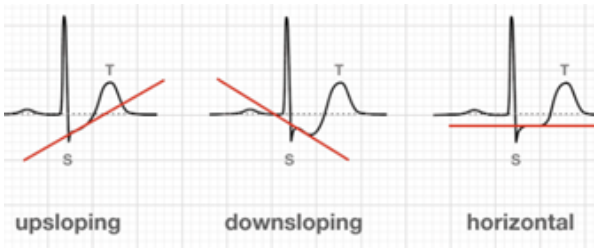
Fasting blood sugar over 120 mg/dL indicates possible diabetes.

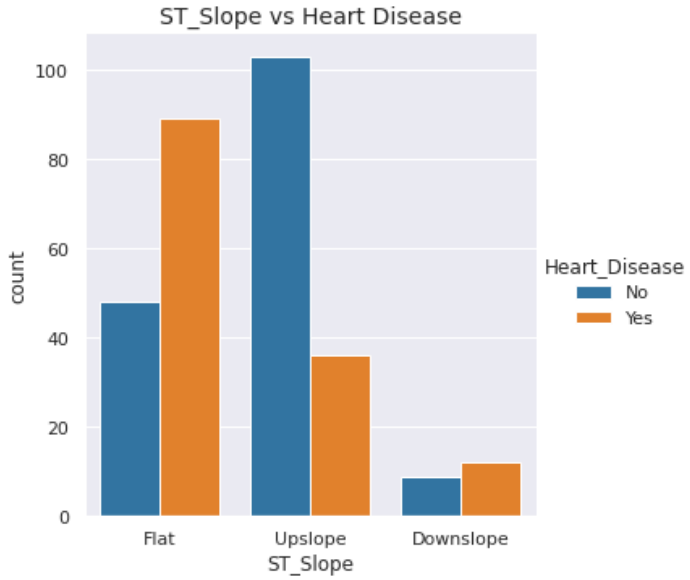
Diabetes is associated with increased risk of heart disease, so one would expect a diabetes marker like this to correlate well with it. In our dataset this relationship turned out to be negligible.

ST_Slope



This is another ECG wave observation





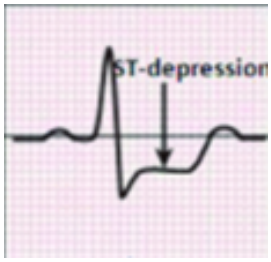
Upslope is the normal, healthy shape

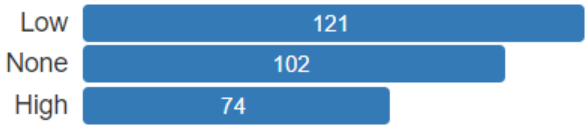
Our Downslope does not show as high a correlation as Flat, but this could be the result of our small dataset, and within that a tiny sample of Downslope records.

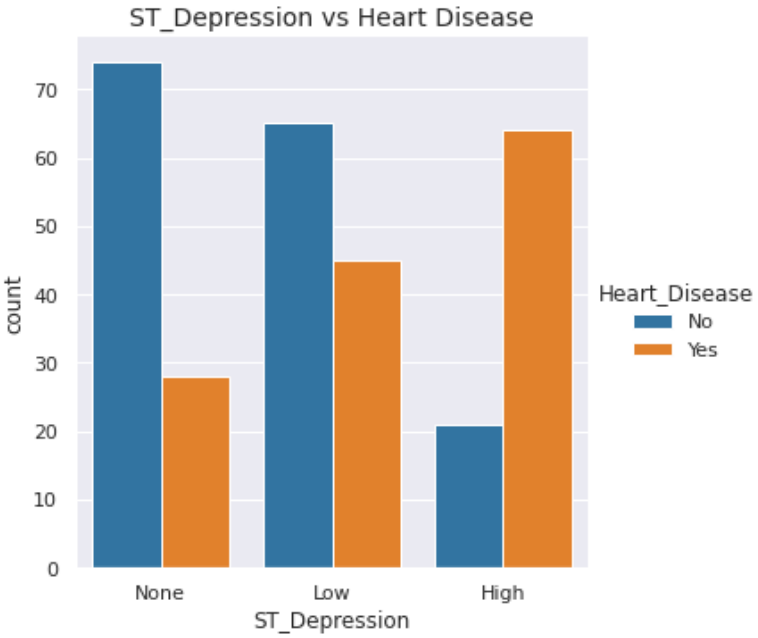
ST_Depression

A second observation of the ST segment in the ECG wave.

ST depression induced by exercise, relative to rest. This was a numeric feature in the original data, we divided it into three ranges, as described above in the numeric data section.



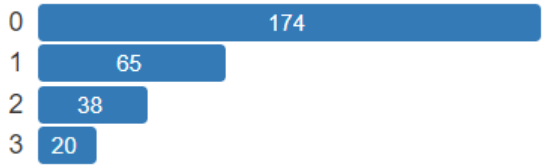


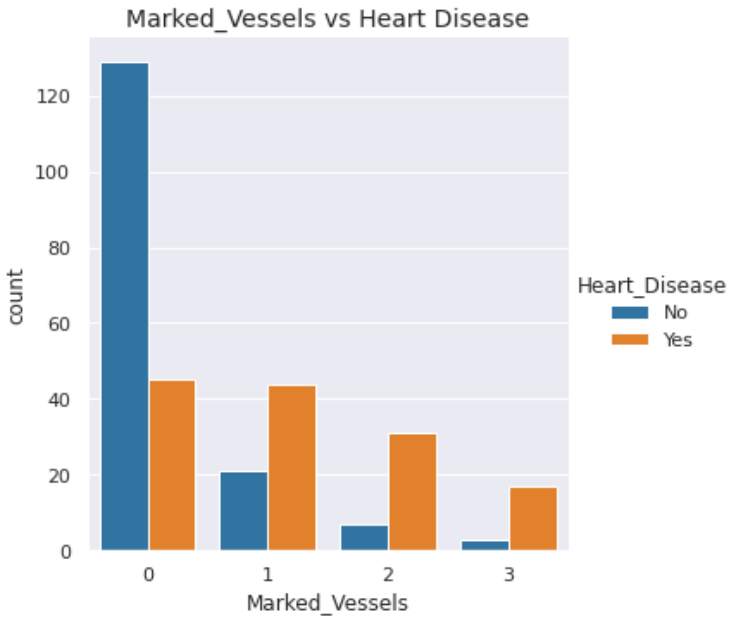


Marked_Vessels

Number of major vessels (0-3) colored by fluoroscopy.

Radioactive dye is introduced to the body followed by x-ray imaging to detect any structural abnormalities present in the heart. The quantity of vessels colored is positively correlated with presence of heart disease.





Data Preprocessing

The data exploration revealed a number of factors that could be adjusted for in the setup of the transformation pipeline to help tune the model.

Sample and Split

Train Test Split - The default model train/test split of 70% (0.7) did not perform well due to the smaller size of the data set. To provide more data for the model training, the train size was increased to 90% (0.9)

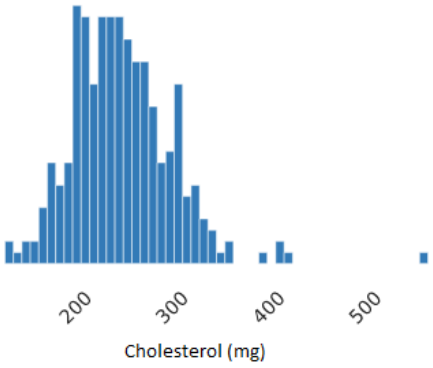
```
train_size = 0.9
```

Data Preparation

Ordinal Encoding - By binning the ST_Depression column into “None”, “Low”, “High” ordinal encoding was added during setup to retain their intrinsic order/ranking.

```
ordinal_features = {'ST_Depression': ['None','Low','High']}
```

Outliers - During data exploration anomalous values were encountered in a number of records. For this reason remove_outliers was set to True during the setup. For example the column cholesterol has a mean of 247, standard deviation of 52, and a few numbers that are 5+ standard deviations from the mean. This is not a long tail, but sporadic data in a sparse tail.



Value	Count	Frequency (%)
564	1	0.3%
417	1	0.3%
409	1	0.3%
407	1	0.3%

With an initial set of 267 records, outlier detection removed 12 down to 255 records, with a Train/Test split of 228/27.

```
remove_outliers = True
```

Scale and Transform

Normalization - The dataset has several numeric columns, each with a very broad range of possibilities. Normalization was applied to rescale the values of the numeric columns.

```
normalize = True
```

Feature Engineering

Bin Numeric Features - Due to the number and range of unique values , the continuous value of “Age” was binned during setup to minimize influence on the trained model

```
bin_numeric_features = ['Age']
```

Feature Selection

Remove Multicollinearity - Data analysis shows that a number of columns possess a high correlation. To minimize this impact on the model “Remove Multicollinearity” was set to true.

```
remove_multicollinearity = True
```

Principle Component Analysis - Due to high level of related data PCA was set to True during setup to reduce the dimensionality of the data

```
pca = True
```

Model Comparison

When generating the initial comparison of models several options were utilized:

Number of Folds - Number of folds was explored and left at the default value of 10. Due to the smaller size of the data set a reduction in the number of folds would reduce the size of the training component on the "train/test" split during model creation. Maximizing the amount of data in the train split provided a positive benefit to the model.

Optimization - There are three different directions to tune a classification model, depending on what you are trying to achieve and how it's reflected in the Confusion Matrix. If the goal is to maximize True predictions, then Accuracy is the only metric needed, because all Trues are good. On the other hand, with False predictions we often prefer to err on one side or the other, depending on the relative cost of an incorrect Positive vs an incorrect Negative prediction. Precision is how we minimize false Positives, and Recall is minimizing false Negatives. In our situation we don't want to miss an actual case of disease, so we tune our model to optimize its Recall performance.

Additional Models - The compare model feature of "Turbo" was set to false to enable the inclusion of additional models in our comparison. The additional models are more resource intensive in their creation and so are left out by default.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
mlp	MLP Classifier	0.7976	0.9052	0.7845	0.8042	0.7792	0.5944	0.6125	0.564
gbc	Gradient Boosting Classifier	0.8024	0.8819	0.7845	0.8079	0.7829	0.6037	0.6204	0.147
rbfsvm	SVM - Radial Kernel	0.8152	0.8979	0.7836	0.8261	0.7941	0.6282	0.6422	0.023
lightgbm	Light Gradient Boosting Machine	0.8148	0.8828	0.7818	0.8219	0.7938	0.6267	0.6367	0.074
lr	Logistic Regression	0.8067	0.9007	0.7655	0.8226	0.7809	0.6104	0.6251	0.266
et	Extra Trees Classifier	0.8026	0.8949	0.7645	0.8130	0.7731	0.6009	0.6160	0.464
qda	Quadratic Discriminant Analysis	0.7761	0.8443	0.7564	0.7761	0.7550	0.5509	0.5635	0.016
knn	K Neighbors Classifier	0.8283	0.8738	0.7555	0.8676	0.8015	0.6526	0.6638	0.118
ridge	Ridge Classifier	0.8111	0.0000	0.7373	0.8485	0.7772	0.6172	0.6329	0.015
lda	Linear Discriminant Analysis	0.8111	0.8949	0.7373	0.8485	0.7772	0.6172	0.6329	0.017
dt	Decision Tree Classifier	0.7370	0.7373	0.7355	0.7101	0.7189	0.4720	0.4770	0.017
rf	Random Forest Classifier	0.7897	0.8760	0.7273	0.8223	0.7585	0.5752	0.5919	0.494
svm	SVM - Linear Kernel	0.7583	0.0000	0.7173	0.7787	0.7326	0.5138	0.5287	0.014
nb	Naive Bayes	0.7634	0.8548	0.7091	0.7826	0.7300	0.5228	0.5372	0.016
gpc	Gaussian Process Classifier	0.7974	0.8948	0.7064	0.8360	0.7598	0.5883	0.6003	0.049
ada	Ada Boost Classifier	0.7107	0.7932	0.6909	0.6914	0.6776	0.4181	0.4298	0.109

Model Analysis

Based on the pipeline setup parameters and the options in the compare model, the top performing model while optimizing for Recall is the MLP Classifier (mlp) model.

```
model1 = create_model('mlp')
display(model1)
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7826	0.9615	0.9000	0.6923	0.7826	0.5725	0.5923
1	0.8261	0.8538	0.7000	0.8750	0.7778	0.6378	0.6485
2	0.9565	0.9462	1.0000	0.9091	0.9524	0.9125	0.9161
3	0.7826	0.8636	0.6364	0.8750	0.7368	0.5594	0.5800
4	0.7391	0.8485	0.6364	0.7778	0.7000	0.4733	0.4808
5	0.7826	0.9773	0.5455	1.0000	0.7059	0.5560	0.6205
6	0.6522	0.8258	0.7273	0.6154	0.6667	0.3083	0.3130
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0.6364	0.7917	0.7000	0.5833	0.6364	0.2787	0.2833
9	0.8182	0.9833	1.0000	0.7143	0.8333	0.6452	0.6901
Mean	0.7976	0.9052	0.7845	0.8042	0.7792	0.5944	0.6125
SD	0.1089	0.0721	0.1647	0.1426	0.1132	0.2167	0.2160

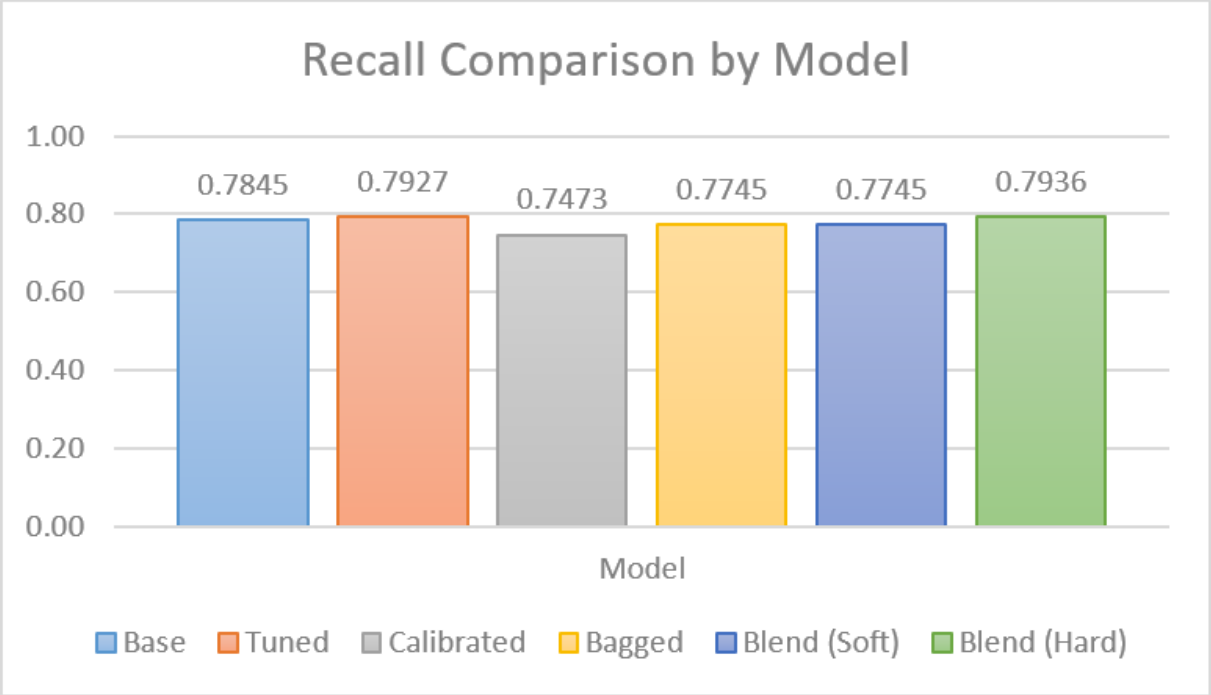
```
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=(100,), learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=500,
              momentum=0.9, 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)
```

Using the MLP Classifier as our base model, analysis and model exploration was completed. Model options were created for:

- Base model
- Tuned model
- Calibrated model
- Bagged model
- Blended model (soft) - (MLP Classifier, Gradient Boosting Classifier, SVM - Radial Kernel)
- Blended model (hard) - (MLP Classifier, Gradient Boosting Classifier, SVM - Radial Kernel)

Model Comparison

Model		Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Base	Mean	0.7976	0.9052	0.7845	0.8042	0.7792	0.5944	0.6125
	SD	0.1089	0.0721	0.1647	0.1426	0.1132	0.2167	0.2160
Tuned	Mean	0.8328	0.9077	0.7927	0.8542	0.8103	0.6629	0.6783
	SD	0.0721	0.0590	0.1355	0.1152	0.0849	0.1454	0.1432
Calibrated	Mean	0.7980	0.8884	0.7473	0.8223	0.7681	0.5922	0.6091
	SD	0.0859	0.0780	0.1624	0.1096	0.0992	0.1725	0.1684
Bagged	Mean	0.8200	0.8994	0.7745	0.8401	0.7943	0.6365	0.6516
	SD	0.0774	0.0631	0.1485	0.1054	0.0940	0.1565	0.1535
Blend (Soft)	Mean	0.8152	0.9108	0.7745	0.8425	0.7933	0.6280	0.6472
	SD	0.0720	0.0588	0.1308	0.1300	0.0786	0.1436	0.1451
Blend (Hard)	Mean	0.8194	0.0000	0.7936	0.8379	0.8020	0.6374	0.6561
	SD	0.0883	0.0000	0.1362	0.1392	0.0906	0.1756	0.1755



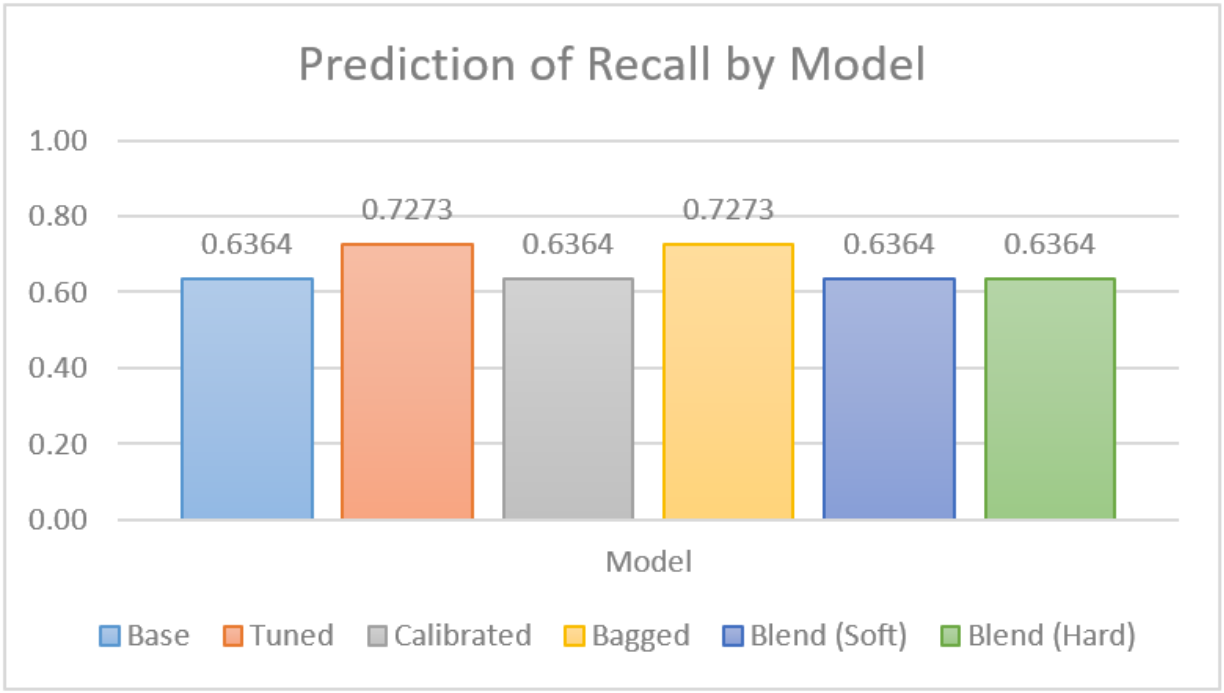
As we are optimizing for Recall, from the chart it's clear that the "Blended model (hard)" at 0.7396 outperforms the next best model, "Tuned" at 0.7927. It is however an extremely marginal increase, and if you include the lower standard deviation of the Tuned model results the difference is negligible.

Model Prediction Comparison

Each model was run through “predict_model” for a further comparison. e.g for tuned_model:

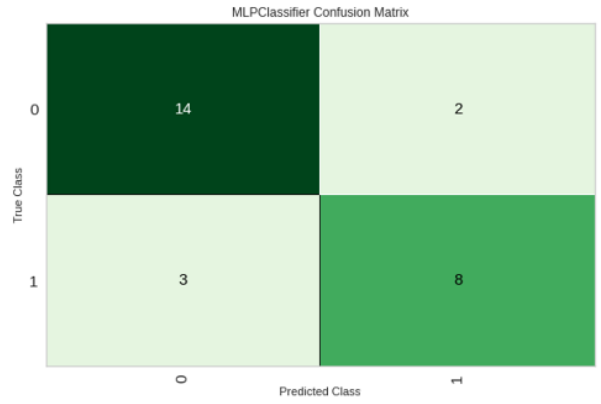
```
results_tuned_model1 = predict_model(tuned_model1)
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Base	**MLP Classifier**	0.7407	0.8750	0.6364	0.7000	0.6667	0.4553	0.4567
Tuned	**MLP Classifier**	0.8148	0.8750	0.7273	0.8000	0.7619	0.6110	0.6128
Calibrated	**MLP Classifier**	0.7778	0.8239	0.6364	0.7778	0.7000	0.5263	0.5330
Bagged	**MLP Classifier**	0.7778	0.8523	0.7273	0.7273	0.7273	0.5398	0.5398
Blend (Soft)	**Voting Classifier**	0.7778	0.8580	0.6364	0.7778	0.7000	0.5263	0.5330
Blend (Hard)	**Voting Classifier**	0.8148	0.7869	0.6364	0.8750	0.7368	0.5994	0.6175

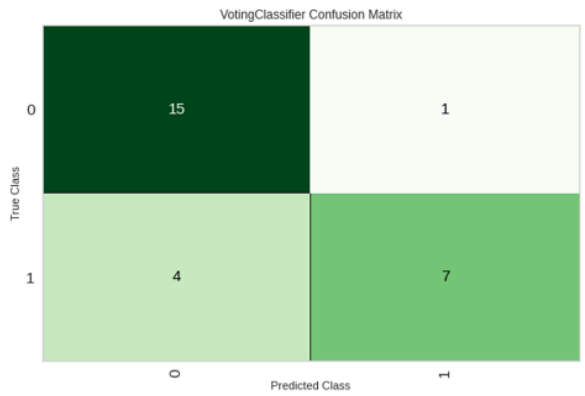


We notice that the Recall results fluctuate between two distinct values 0.6364 and 0.7273. While this difference may seem significant it's important to look at the origin of those numbers. If we examine the confusion matrix for the "Tuned" and the "Blend (Hard)" models:

Tuned Model



Blended Model (Hard)



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

Recall: 8 / (8 + 3) = 0.7273

Recall: 7 / (7 + 4) = 0.6364

Due to the smaller data set size, the recall range on test data is shifted by .0909 from a single observation changing it's prediction category. While the prediction is effective, it's clear that both the model training and the model choice would benefit greatly from a much larger dataset.

Best Model

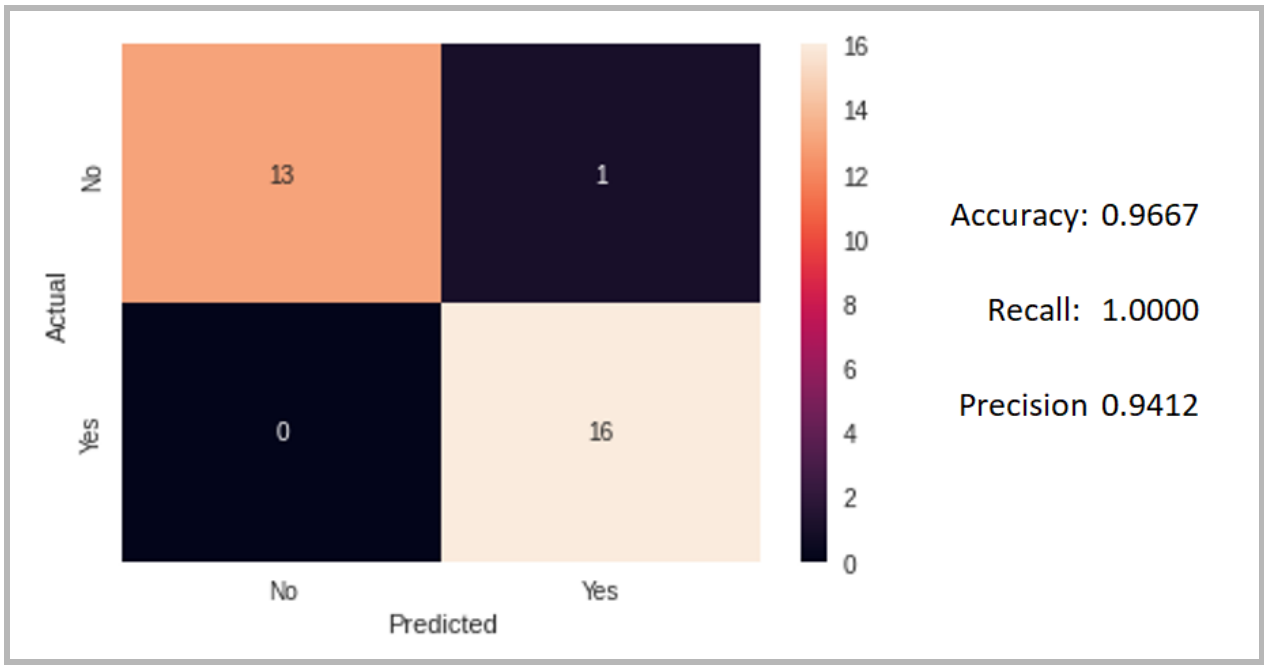
While the model with the best performing “Recall” is the “Blended model (hard)”, the model only outperforms the “Tuned” model by only 0.0009. Additionally the standard deviation in the “Tuned” model is lower by 0.0007.

The more important analysis factor is the smaller size of the data. Since these two models are virtually identical in “Recall” metrics they can be viewed as approximate equivalents. However, since the data size is relatively small, choosing the single “Tuned” MLP model, rather than a more complicated blending of models is a simpler and more applicable approach.

```
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)
```

A prediction of the finalized model against the unseen data was completed with promising results, and no false negatives.

```
results_final_model = predict_model(final_model, data=data_unseen)
```



Conclusion

The prediction modeling provides a satisfactory method for detecting heart disease based on our test results. The current model set is smaller than would be ideal but the model creation, predictions, and testing proves that this would be an effective endeavor. Implementation should include a continuous retraining and update of the model as new information is available. There are a number of opportunities to utilize this model or modified models to improve client care, save money, and improve resource planning.

An area of opportunity would be to conduct a business flow analysis, to evaluate the order in which tests should be completed, and results returned. This could allow opportunities to improve patient triage by scheduling appointments for those who have the higher risk of heart disease. Improvements to client care, budgeting, as well as space and resource planning could be forecast if some modeling predictions can be obtained at earlier decision points.

The model meets our performance criteria with a false-negative rate < 5% and a false-positive rate < 30%, strong model documentation to ensure reliability and repeatability, and a web interface provides simple cross platform usability.

Integrating this model into the business workflow in a more automated fashion would be far more beneficial than the web application. Ideally, as patient test results are entered into an electronic medical system (EMS) it would trigger an API call to the model and flag the patient status. This “flag” could be aggregated across all clients to remove private information, and shared with other stakeholders to improve their business forecasting capabilities to have an idea of the number of patients requiring future support and resources.

The implementation and utilization of this model could have a significant benefit to a number of stakeholders whether they be the patient, doctor, or medical administration. There are two main areas of concern with implementing a model such as this: predicting the client has no heart disease but actually does (false negative), and the transmission/storage/usage of personal information.

To alleviate these concerns two things should occur. First, the business process should be investigated and adjusted to ensure that when the model is integrated into the workflow, adequate rigour is taken to ensure patients do not go undiagnosed. Secondly, the model does not require any personally identifying information, so any API implementation should restrict the collection, submission, transmission of any personal information. Any automated submission and matching of information should ensure that all personal data is obfuscated, and encrypted during transmission as well as at rest.

Bibliography

Health Care Services—Nunavut. 2017 March Report of the Auditor General of Canada. URL: https://www.oag-bvg.gc.ca/internet/english/nun_201703_e_41998.html

Nillsf. Confusion Matrix, accuracy, recall, precision, false positive rate and F-scores explained. URL: https://blog.nillsf.com/index.php/2020/05/23/confusion-matrix-accuracy-recall-precision-false-positive-rate-and-f-scores-explained/

Heart Disease UCI. Kaggle. URL: https://www.kaggle.com/cherngs/heart-disease-cleveland-uci

O. Pelivan. HeartDisease.URL:https://www.kaggle.com/onatto/predicting-heart-disease-a-detailed-guide

Jesse Charis. JCharis Tech. How to Split Dataset into Training and Testing Dataset for Machine Learning. URL:https://blog.jcharistech.com/2020/09/23/how-to-split-dataset-into-training-and-testing-dataset-for-machine-learning/

Pycaret. Pycaret - Preprocessing. URL: https://pycaret.org/preprocessing/

Towards Data Science. Moez Ali. Build and deploy your first machine learning web app. URL: https://towardsdatascience.com/build-and-deploy-your-first-machine-learning-web-app-e020db344a99