Data Science Final Project Presentation

Class: IBM IMVAI-2102

Team: 03

Team Members:
Peh Lay Hock, Teo Kai Heng, Tan Gaik Neo Ivy & Seah Ghim Sin Steven

Don’t Ignore the Pain

Using Data Science to Predict of Coronary Heart Disease (CHD) and Helping Healthcare Professionals to Make Better Decision on CHD Testing

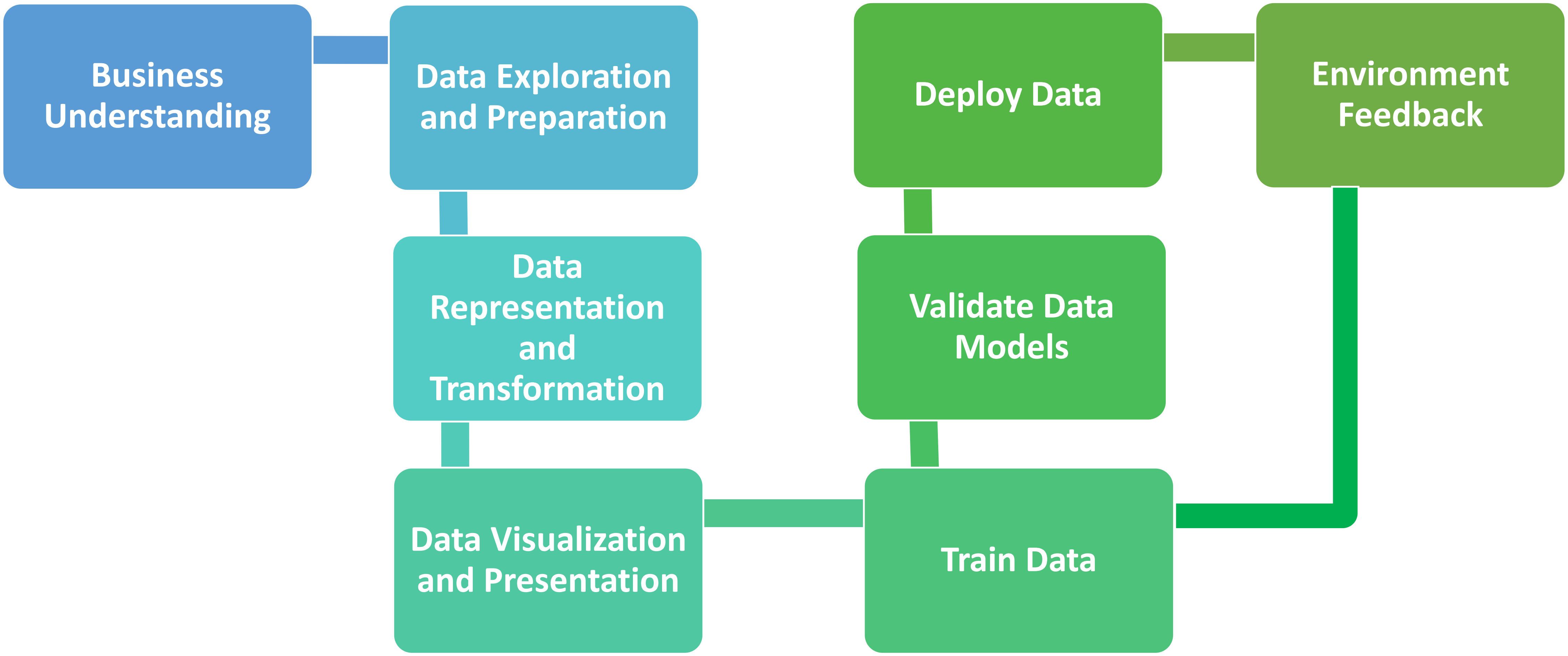
Introduction

Presentation Flow & Team Roles

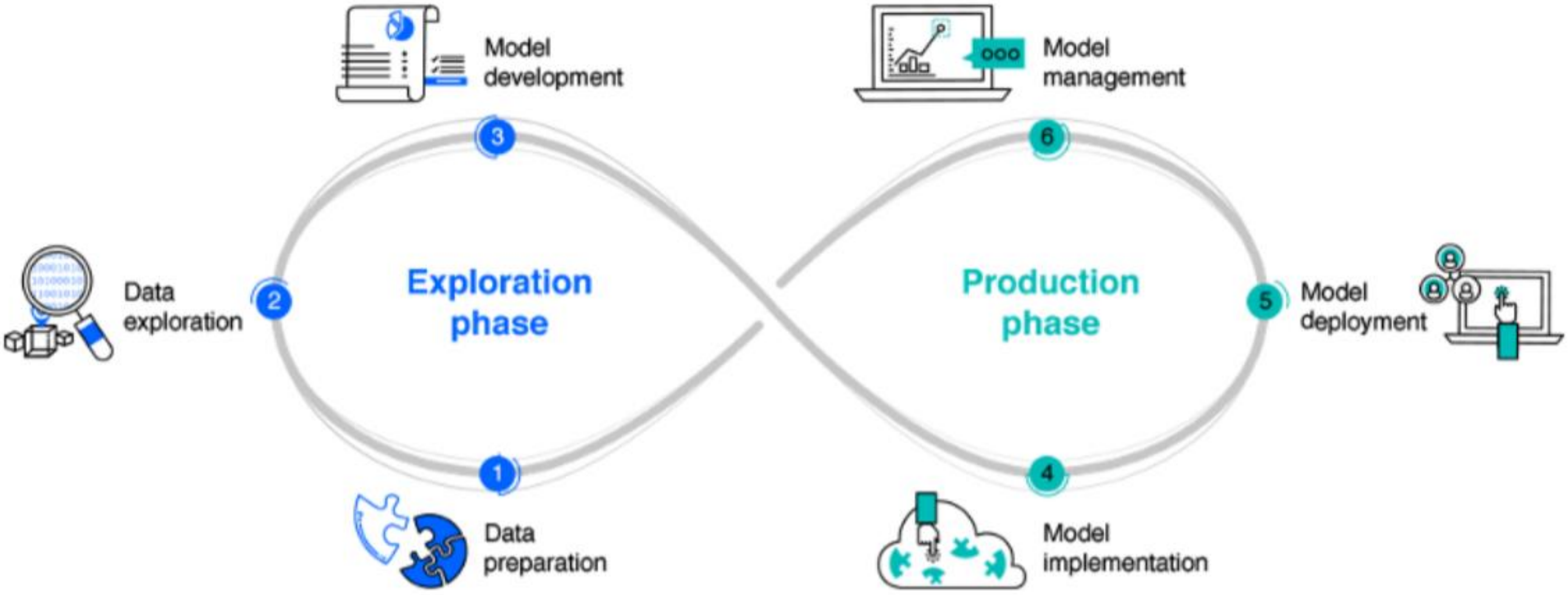
Name	Topics
Steven Seah	Introduction Planning Analytics
Ivy Tan	Descriptive Analytics Diagnostic Analytics
Teo Kai Heng	Predictive Analytics
Peh Lay Hock	Prescriptive Analytics Summary and Reflection

Data Analytics Lifecycle

Data Science Methodology



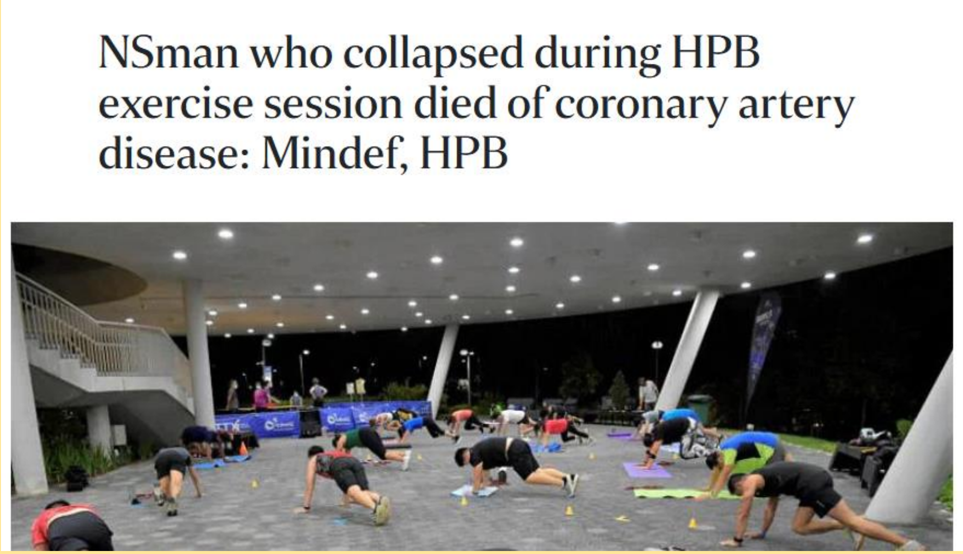
Data Science Lifecycle



Planning Analytics

Project Background

Coronary Heart Disease (CHD) presents in two main forms: myocardial infarction (heart attack) and angina. Coronary heart disease is now the leading cause of death worldwide. An estimated 3.8 million men and 3.4 million women die each year from CHD


NSman who collapsed during HPB exercise session died of coronary artery disease: Mindef, HPB


NEWS SPORTS ENTERTAINMENT LIFESTYLE RACING

SINGAPORE

NUS professor, 53, who volunteered with police force, dies following afternoon run

Business Understanding

What is the business problem?

Complaints are up recently especially in the situation of patients not getting fast enough diagnosis for CHD during admission in A&E.

In some cases, patients were discharged with no conclusive diagnosis for CHD and later found to actually have CHD. This ultimately led to hospital’s reputation at stake in the area of healthcare providers’ competency

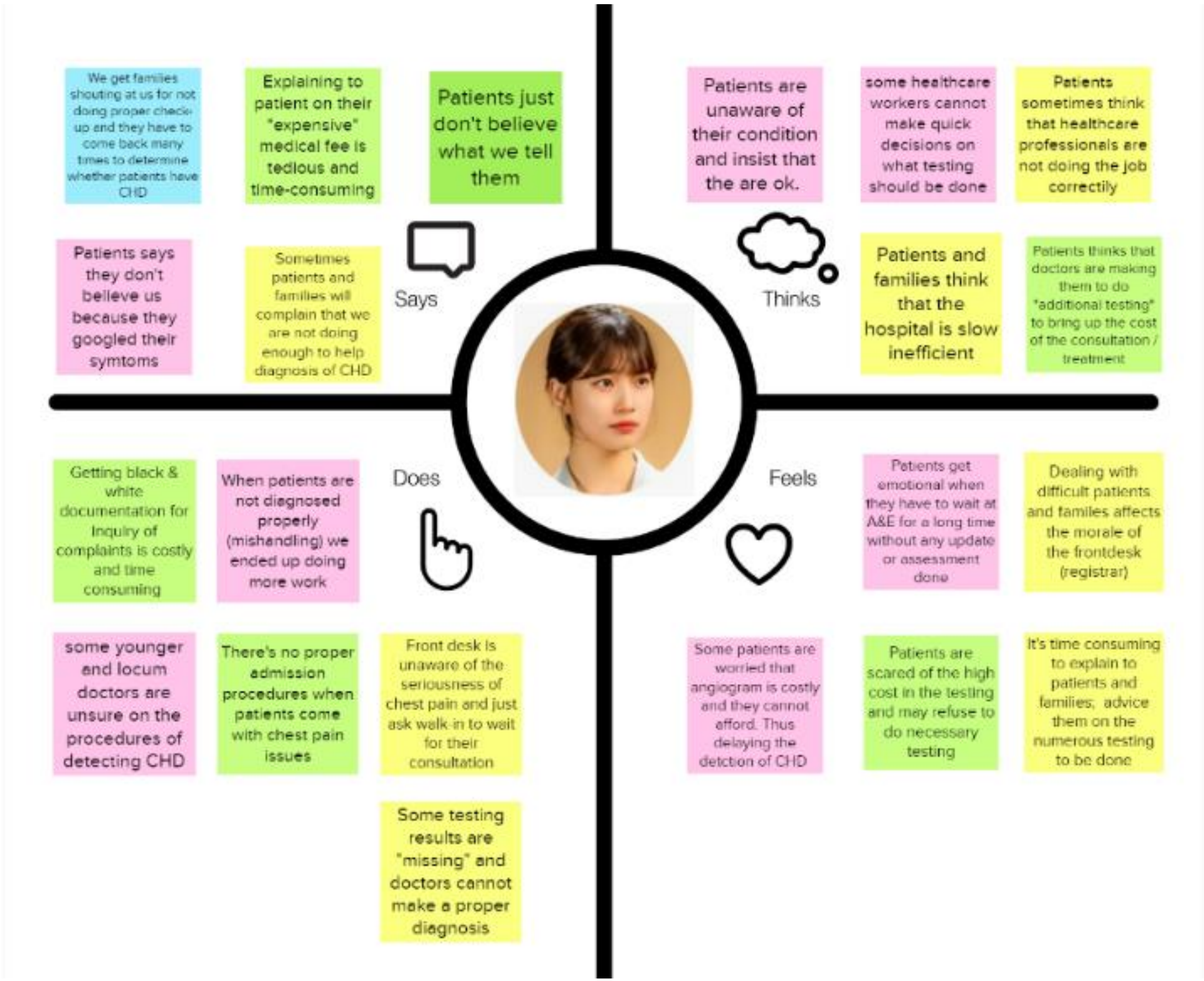
What is the business opportunity?

- Help the Hospital Administrator to response to patients' queries and complaints quickly and with more transparency.
- Failsafe checklist of patients' symptoms and improve the processes for testing on patients who are suspected with CHD
- Avoid time wasting at the initial diagnosis process during admission
- Reduce miscommunication between patients and healthcare professionals with regards to diagnosis and billing

Our Persona

Name:	Ms. Hope Foo
Age:	35
Education:	Degree
Job:	Hospital Administrator
Work Experience:	10 years

Empathy Map



We get families shouting at us for not doing proper check-up and they have to come back many times to determine whether patients have CHD
Explaining to patient on their "expensive" medical fee is tedious and time-consuming
Patients just don't believe what we tell them
Patients says they don't believe us because they googled their symtoms
Sometimes patients and families will complain that we are not doing enough to help diagnosis of CHD
Says
Patients are unaware of their condition and insist that the are ok.
some healthcare workers cannot make quick decisions on what testing should be done
Patients sometimes think that healthcare professionals are not doing the job correctly
Thinks
Patients and families think that the hospital is slow inefficient
Patients thinks that doctors are making them to do "additional testing" to bring up the cost of the consultation / treatment
Getting black & white documentation for Inquiry of complaints is costly and time consuming
When patients are not diagnosed properly (mishandling) we ended up doing more work
Does
some younger and locum doctors are unsure on the procedures of detecting CHD
There's no proper admission procedures when patients come with chest pain issues
Front desk is unaware of the seriousness of chest pain and just ask walk-in to wait for their consultation
Some testing results are "missing" and doctors cannot make a proper diagnosis
Feels
Patients get emotional when they have to wait at A&E for a long time without any update or assessment done
Dealing with difficult patients and familes affects the morale of the frontdesk (registrar)
Some patients are worried that angiogram is costly and they cannot afford. Thus delaying the detction of CHD
Patients are scared of the high cost in the testing and may refuse to do necessary testing
It's time consuming to explain to patients and families, advice them on the numerous testing to be done

Motivations / Goals / Needs

Motivations

To reduce the number of complaints epecially from those admitted with CHD

Simplify the process of admission and testing for "suspected" cases of CHD

Helps healthcare professionals to have a proper knowledge and know-how to diagnose CHD

Goals

A checklist on the proper procedure of testing and diagnosis

Reduce the miscommunication between healthcare professionals and patients

Make information about CHD treatment and testing more available to patients and families

Needs

Tool / process to identify CHD quickly

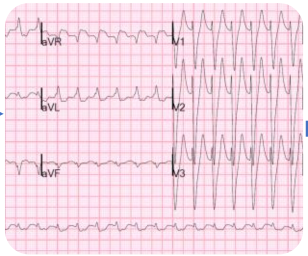
The best indicator(s) of CHD so that healthcare workers can prioritize what test to administer to patients

AS-IS Scenario

Admission with chest pain

Patient waited at lobby

Doctor take BP, Chol and Blood Sugar



Doctor advised to do an ECG

2 hours later

Doctor said everything's OK

Patient / family got angry & asked for more testing

Angiogram confirmed patient has CHD

Pain Points

RESULTING INCREASED COMPLAINTS AND WORKLOAD FOR PERSONA

- Hospital reputation on CHD treatment (patients & families perceptions)
- Getting the result for CHD takes too long
- Confusion when looking at medical report
- Mistrust between patients and healthcare professionals

AFFECTING PATIENTS

- Anxiety from expensive cost for testing and treatment

How Might We Statement

How might we **determine patients' Coronary Heart Disease (CHD) quickly** upon admission**, build trust** through our healthcare **competency** and possibly **SAVE LIVES**

Hypotheses

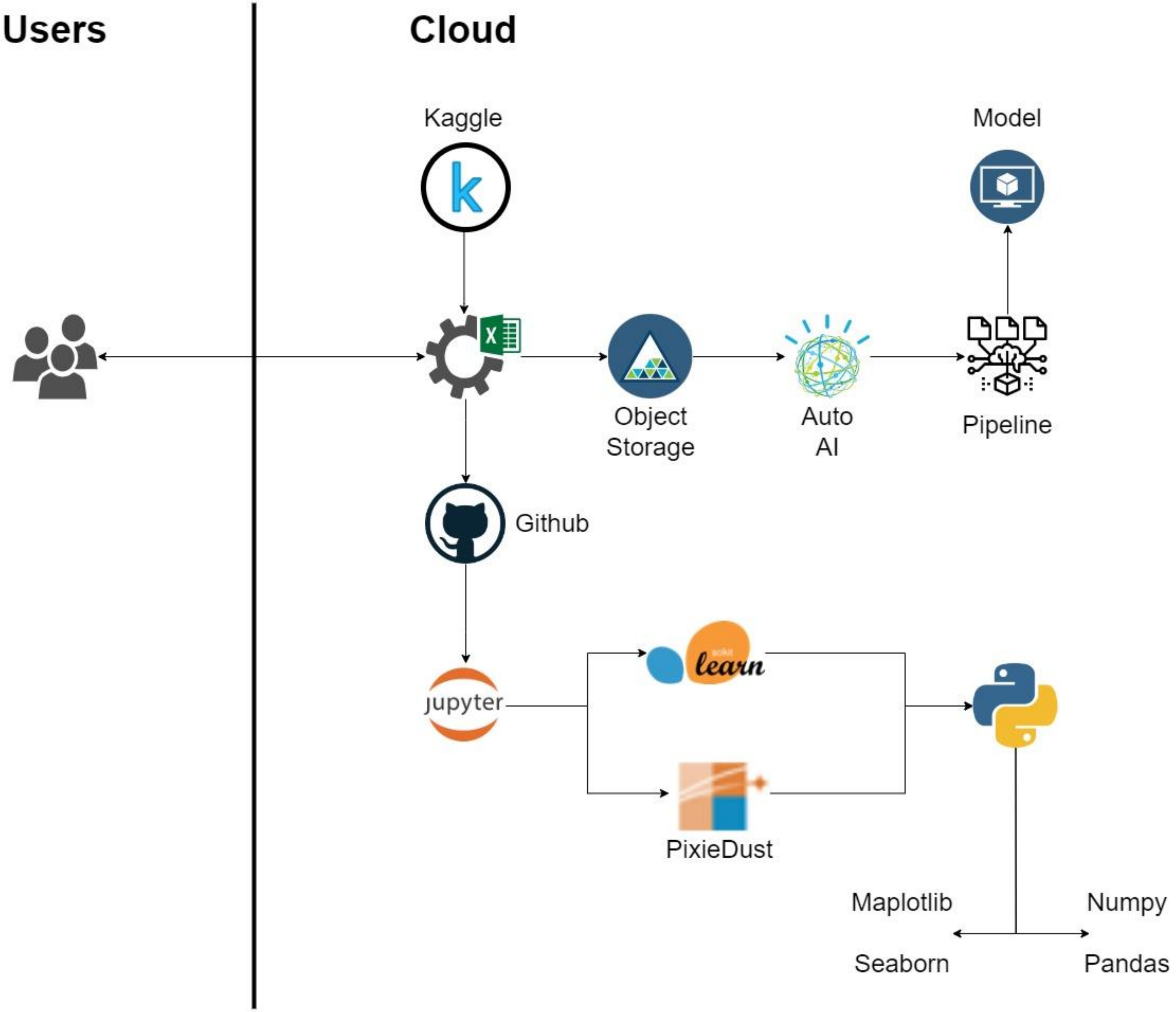
Hypothesis	Statement
H1	Patient admitted with symptoms of chest pain may have Coronary Heart Disease (CHD) depending on the age, gender and rest ECG assessment.
H2	Patients with high levels of blood pressure, blood sugar, cholesterol may likely to have Coronary Heart Disease (CHD)

Descriptive Analytics

Tools

- Mural – for collaboration and Enterprise Design Thinking process
- Excel – for Data Cleaning and Transformation
- IBM Watson Studio – for data analytics
    - Data Visualization
    - Auto AI
    - Machine Learning
    - Cloud Object Storage
- Jupyter Notebook
    - Python
    - Libraries:
        - Pandas
        - Matplotlib
        - Numpy
        - Scikit-Learn
        - PixieDust

Implementation Architecture



Data Exploration

Data Source: UCI Machine Learning Repository

Link: https://archive.ics.uci.edu/ml/datasets/heart+disease

Kaggle: Heart Disease Dataset contributed by David Lapp

Link: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

Creators:

1. Cleveland Clinic Foundation (cleveland.data)
2. Hungarian Institute of Cardiology, Budapest (hungarian.data)
3. V.A. Medical Center, Long Beach, CA (long-beach-va.data)
4. University Hospital, Zurich, Switzerland (switzerland.data)

The Project objective is to build a data model that can predict whether a person diagnosed has heart disease based on patterns extracted from analyzing 14 descriptive features out of total of 76 recorded.

Data Exploration

No	Attribute	Name	Description
1	age	Age	age in years
2	sex	Sex	sex (1 = male; 0 = female)
3	cp	Chest Pain	cp: chest pain type (Values 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4:asymptomatic)
4	trestbps	Rest Blood Pressure	trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5	chol	Serum Cholestorol	chol: serum cholestoral in mg/dl
6	fbs	Fasting Blood Sugar	fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7	restecg	Resting Electrographic Results	restecg: resting electrocardiographic results (Value 0: normal, 1: having ST-T wave abnormality, 2: showing probable or definite left ventricular hypertrophy)
8	thalach	Max Heart Rate	thalach: maximum heart rate achieved
9	exang	Exercise Induced Angina	exang: exercise induced angina (1 = yes; 0 = no)
10	oldpeak	ST Depression	oldpeak = ST depression induced by exercise relative to rest
11	slope	Slope of of the ST Segment	slope: the slope of the peak exercise ST segment (Value 1: upsloping, 2: flat, 3: downsloping)
12	ca	No of vessels	ca: number of major vessels (0-3) colored by flourosopy
13	thal	Heart status from Thallium test	thal: 0 = normal; 1 = fixed defect; 2 = reversable defect
14	target (num)	Classification of the Heart Disease	num: diagnosis of heart disease (angiographic disease status) Value 0: < 50% diameter narrowing (no heart disease) Value 1 to 4 > 50% diameter narrowing (Diff heart diseases)

Data Preparation (Cleansing)

- Data cleansing is not required as the obtained raw data does not have any missing values nor are there any corrupted or incorrectly formatted data

- There is also no outlier observed from the data as the data are found to be coherent, hence none of the data points shall need to be removed due to being outliers.

Data Transformation

cp_True	trestbps_under_130	chol_under_200	chol_bps_fbs_HIGH	max_healthyHR	thalach_under_maxhealthHR
0	1	0	0	168	1
0	0	0	1	167	1
0	0	1	0	150	1

chol_under_200: chol<200 vs chol >=200

max_healthyHR: 220-age

chol_bps_fbs_HIGH: (trestbps_under_120 =0, chol_under_200=0 and fbs =1) vs NOT(trestbps_under_120=0, chol_under_200=0 and fbs =1)

Thalach_under_maxhealthHR: thalach<max_healthyHR vs thalach>=max_healthyHR

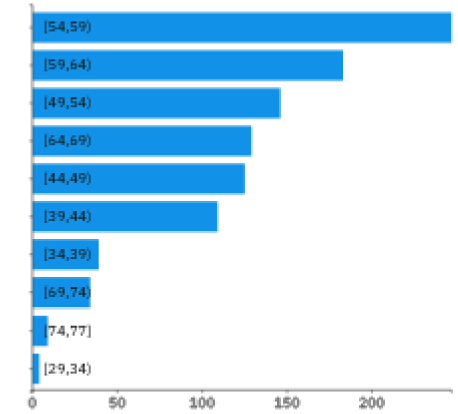
trestbps_under_130: trestbps<130 vs trestbps >=130

Types of CP transformed to Pain (1) or No Pain (0)

Data Visualization (Descriptive statistics of each feature)

age
Integer

FREQUENCY



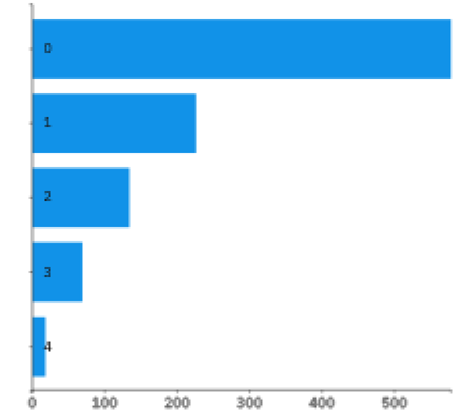
STATISTICS

Interquartile Range	13
Minimum	29
Maximum	77
Median	56
Standard Deviation	9.07229023324428

Age range 29 – 77, Median age is 56 and majority is between 50s – 60s

ca
Integer

FREQUENCY



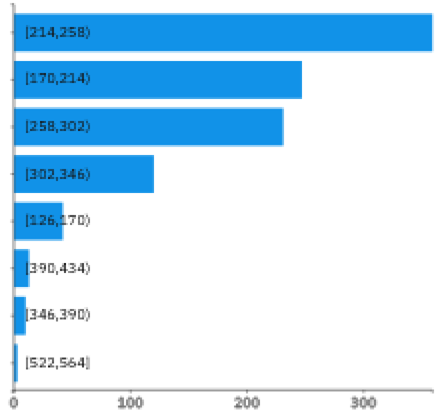
STATISTICS

Interquartile Range	1
Minimum	0
Maximum	4
Median	0
Standard Deviation	1.03079766502428

This feature is not used in our analysis

chol
Integer

FREQUENCY



STATISTICS

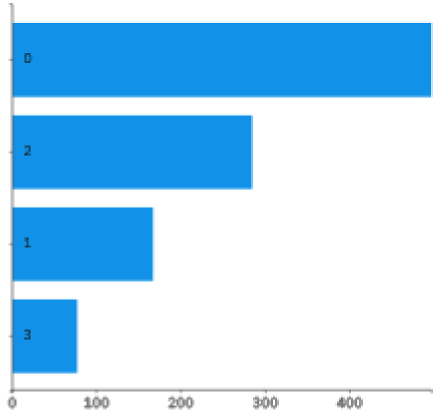
Interquartile Range	64
Minimum	126
Maximum	564
Median	240
Standard Deviation	51.5925102061821

Most subjects have recorded high cholesterol

Data Visualization (Descriptive statistics of each feature)

cp
Integer

FREQUENCY



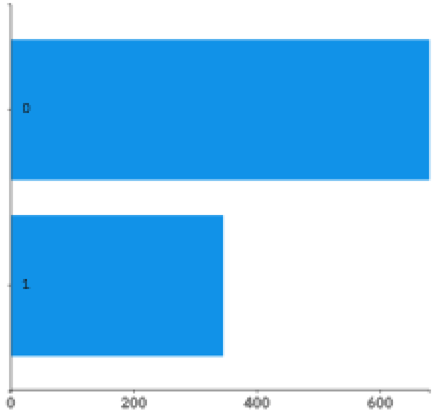
STATISTICS

Interquartile Range	2
Minimum	0
Maximum	3
Median	1
Standard Deviation	1.02964074364586

Chest pain (1 – 3) is fairly distributed with no chest pain (0)

exang
Integer

FREQUENCY



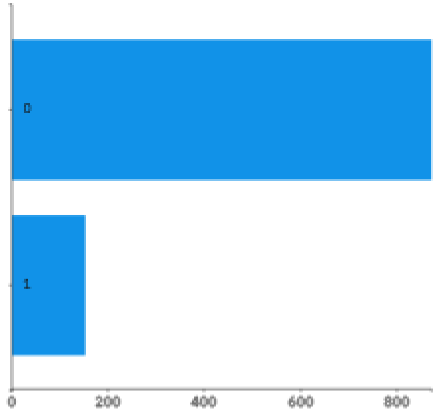
STATISTICS

Interquartile Range	1
Minimum	0
Maximum	1
Median	0
Standard Deviation	0.47277237600371

This feature may be related to CHD

fbs
Integer

FREQUENCY



STATISTICS

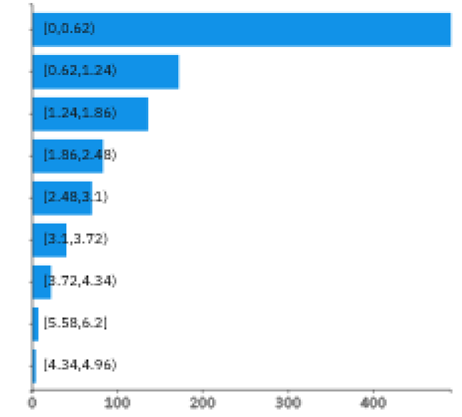
Interquartile Range	0
Minimum	0
Maximum	1
Median	0
Standard Deviation	0.356526689727159

Majority of subjects showed healthy FBS

Data Visualization (Descriptive statistics of each feature)

oldpeak
Decimal

FREQUENCY



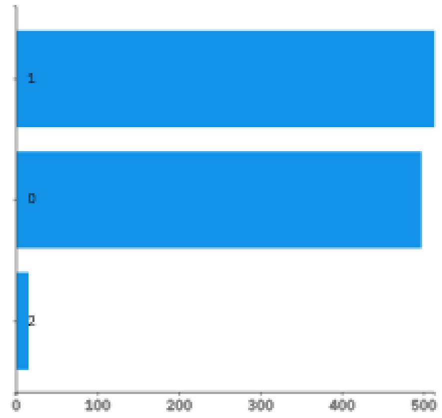
STATISTICS

Interquartile Range	1.8
Minimum	0
Maximum	6.2
Median	0.8
Standard Deviation	1.17505325515017

Subject with lower Old Peak may be linked to CHD

restecg
Integer

FREQUENCY



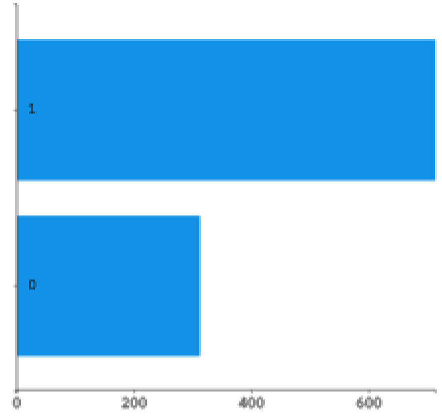
STATISTICS

Interquartile Range	1
Minimum	0
Maximum	2
Median	1
Standard Deviation	0.527877566874893

Subjects with abnormal rest ECG may be linked to CHD

sex
Integer

FREQUENCY



STATISTICS

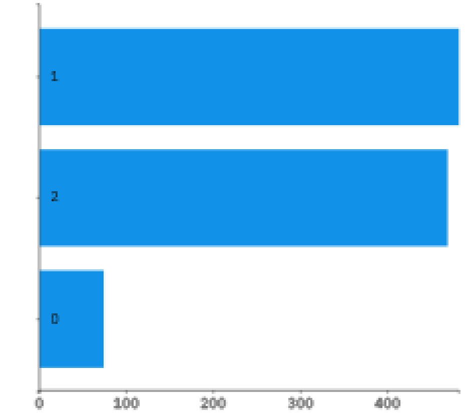
Interquartile Range	1
Minimum	0
Maximum	1
Median	1
Standard Deviation	0.46037332411965

There are more Male than Female test subject. No direct relation to CHD based on gender

Data Visualization (Descriptive statistics of each feature)

slope
Integer

FREQUENCY



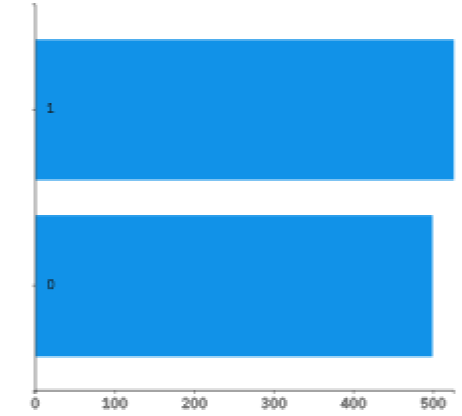
STATISTICS

Interquartile Range	1
Minimum	0
Maximum	2
Median	1
Standard Deviation	0.617755267174591

Subjects with flat and downsloping may be linked to CHD

target
Integer

FREQUENCY



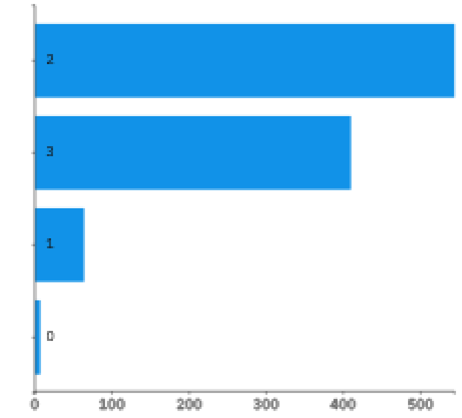
STATISTICS

Interquartile Range	1
Minimum	0
Maximum	1
Median	1
Standard Deviation	0.500070498078805

The sample size is roughly balanced for the different target classes

thal
Integer

FREQUENCY



STATISTICS

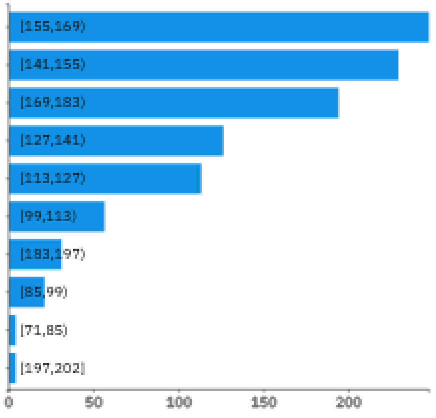
Interquartile Range	1
Minimum	0
Maximum	3
Median	2
Standard Deviation	0.620660238051028

This is an alternative test for CHD

Data Visualization (Descriptive statistics of each feature)

thalach
Integer

FREQUENCY



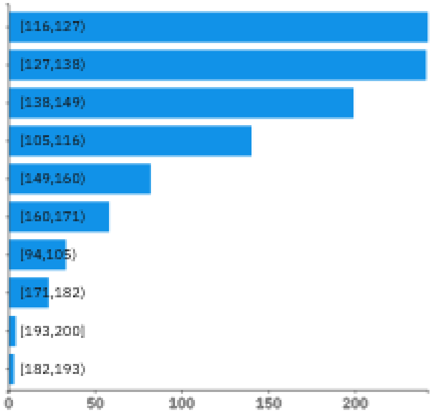
STATISTICS

Interquartile Range	34
Minimum	71
Maximum	202
Median	152
Standard Deviation	23.0057237459772

Subjects with higher max heart rate may be linked to CHD

trestbps
Integer

FREQUENCY

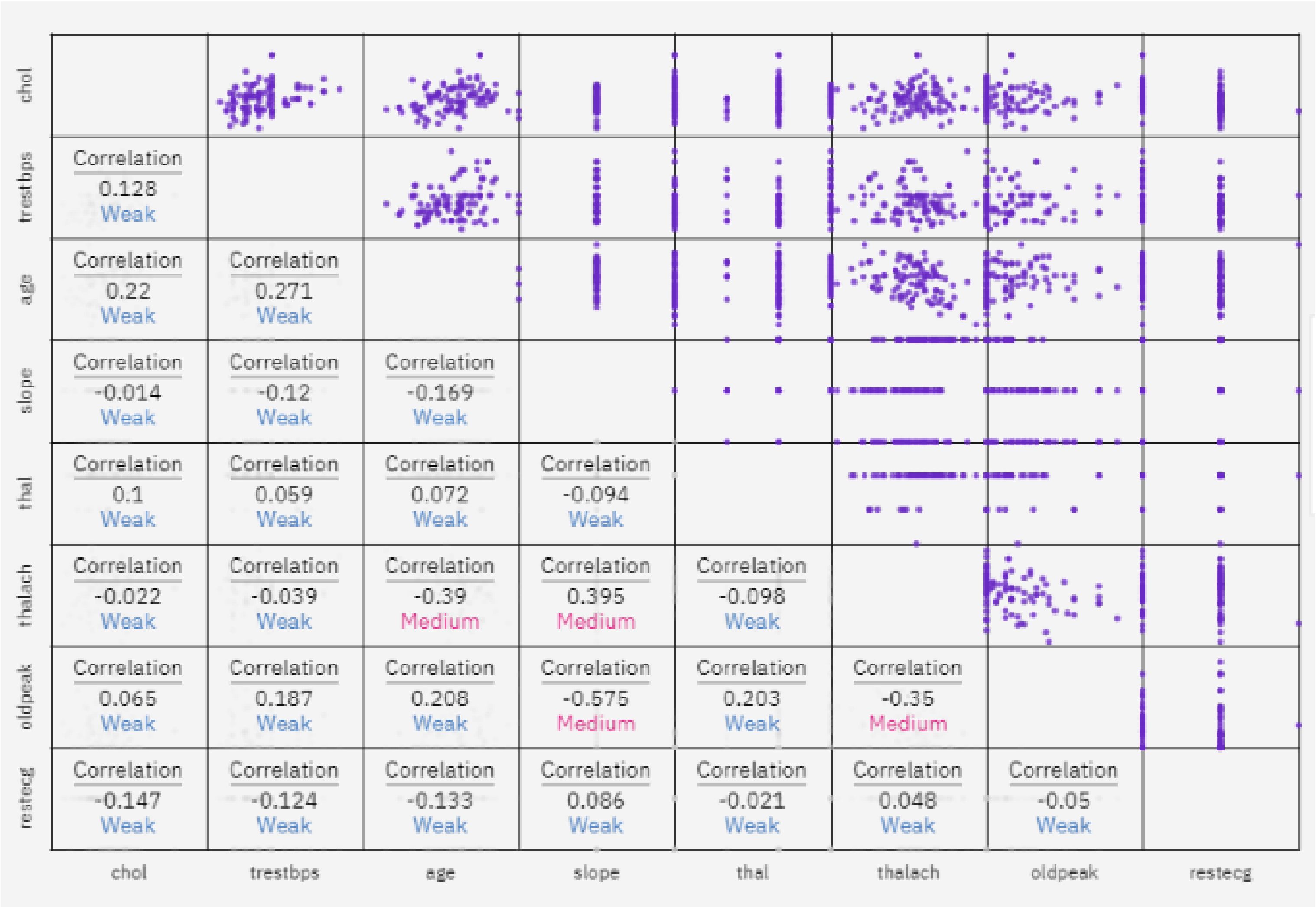


STATISTICS

Interquartile Range	20
Minimum	94
Maximum	200
Median	130
Standard Deviation	17.5167180053764

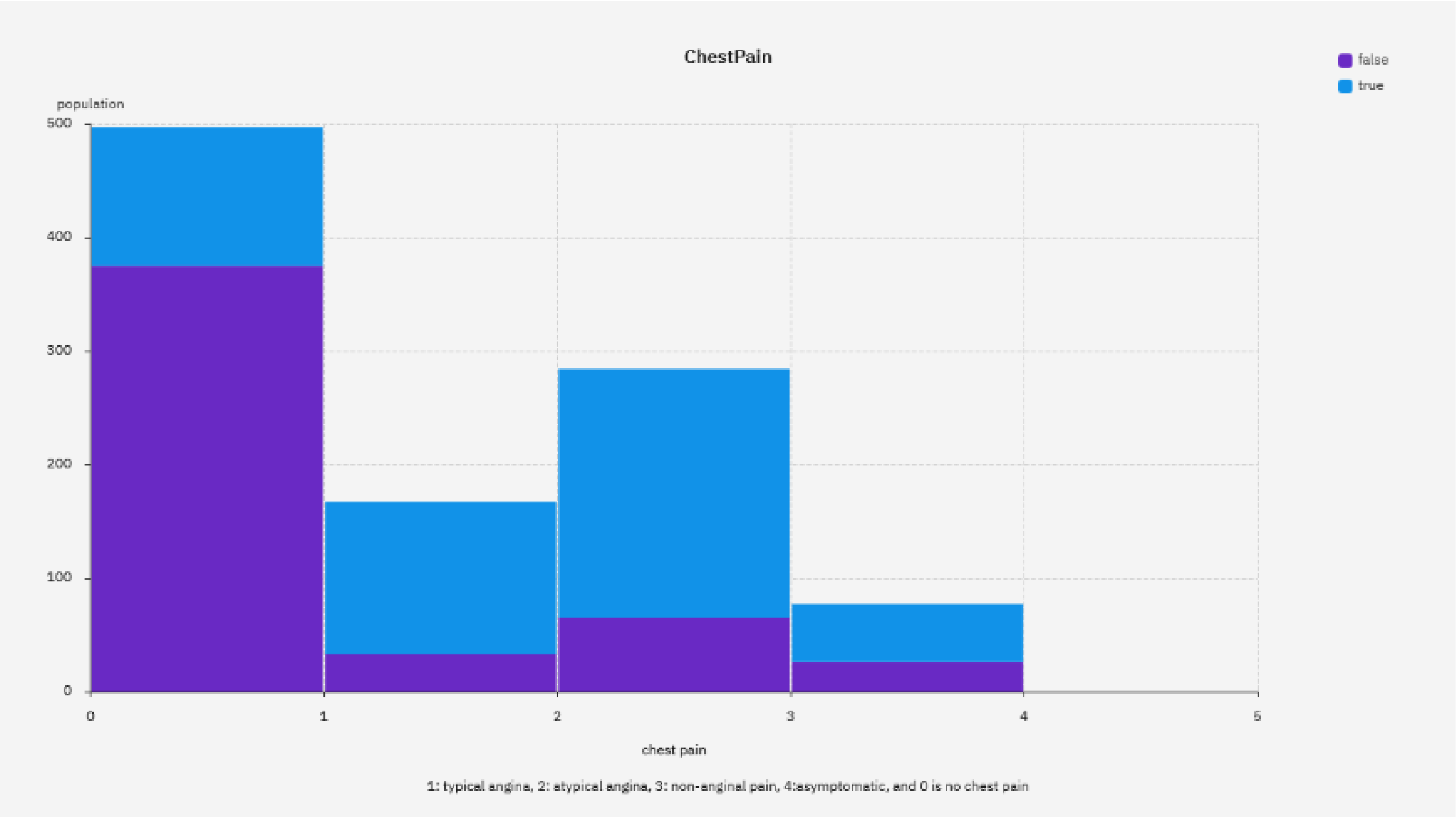
Subjects with high rest blood pressure may be link to CHD

Data Visualization



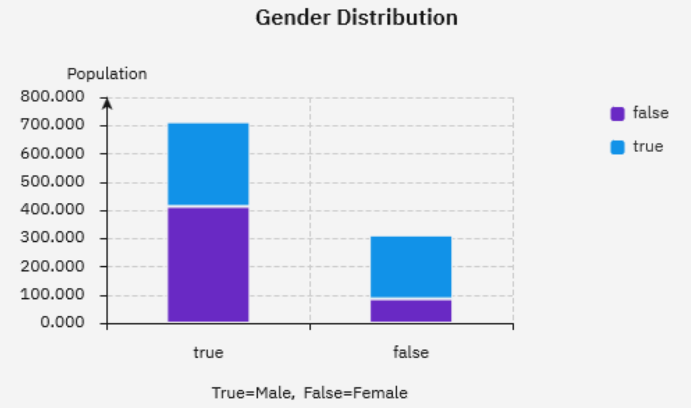
Diagnostic Analytics

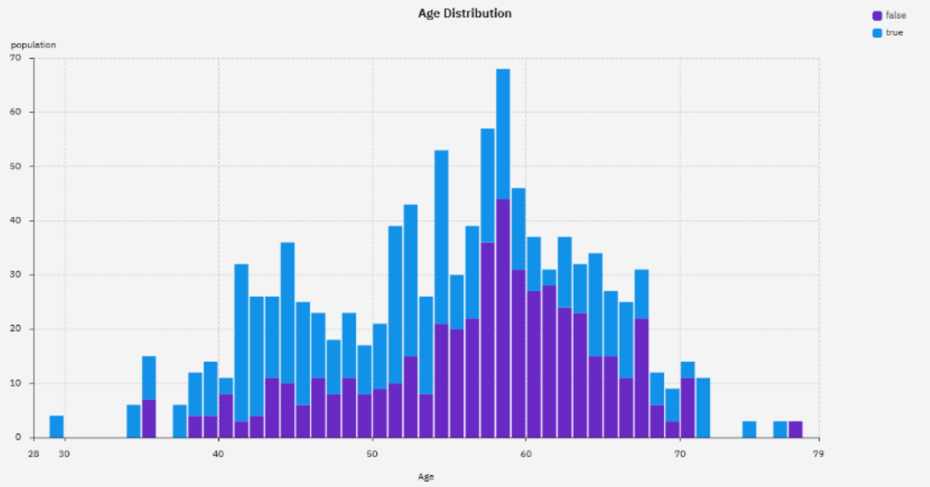
Attribute Insights and Diagnostic Analysis (H1)

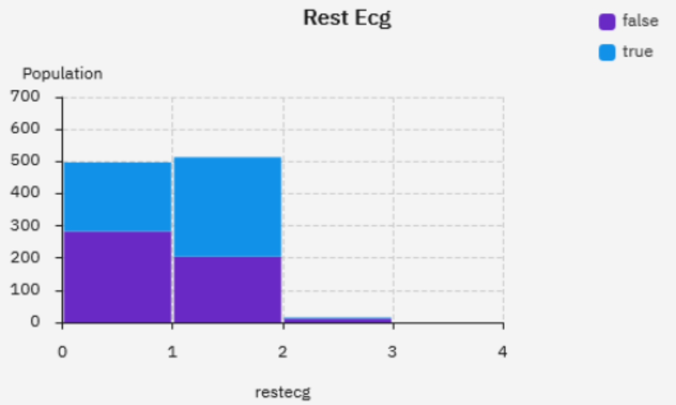


Chest pain seems to be one of the leading indicators for CHD

Attribute Insights and Diagnostic Analysis (H1)

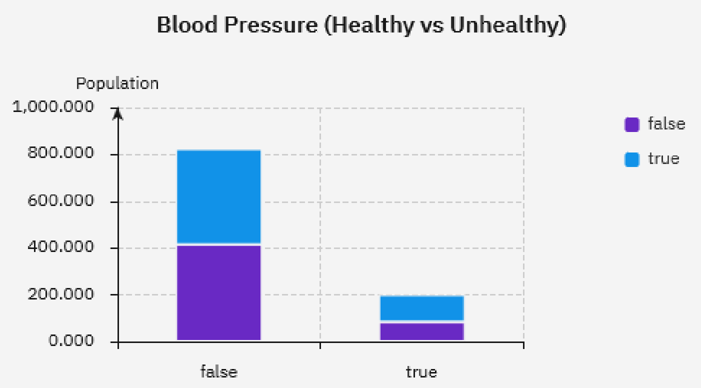




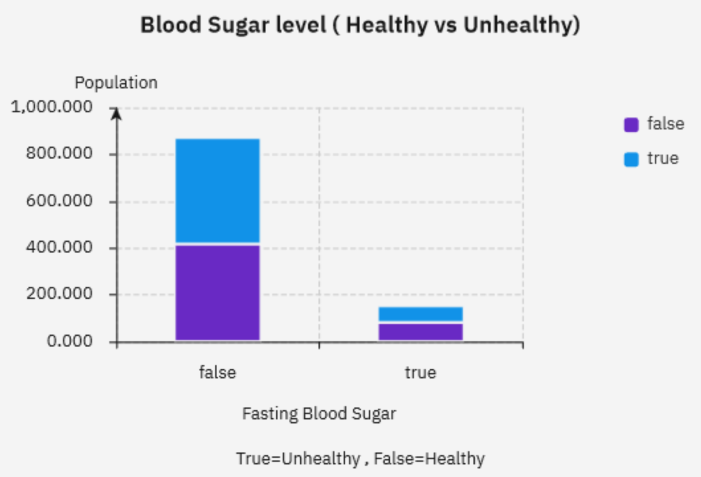


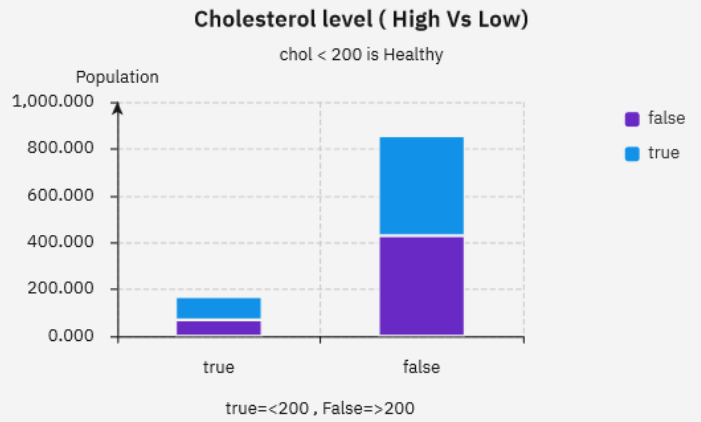
- Age, gender and rest ECG shows trend of being diagnosed with CHD
- Older group, female and abnormal rest ecg seems to indicate higher likelihood of CHD generally.
- There are lesser cases of CHD reported for 30s age group

Attribute Insights and Diagnostic Analysis (H2)



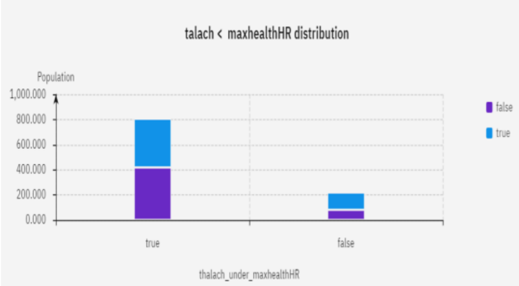
Person’s resting blood pressure < 130
(False = unhealthy and true = healthy)

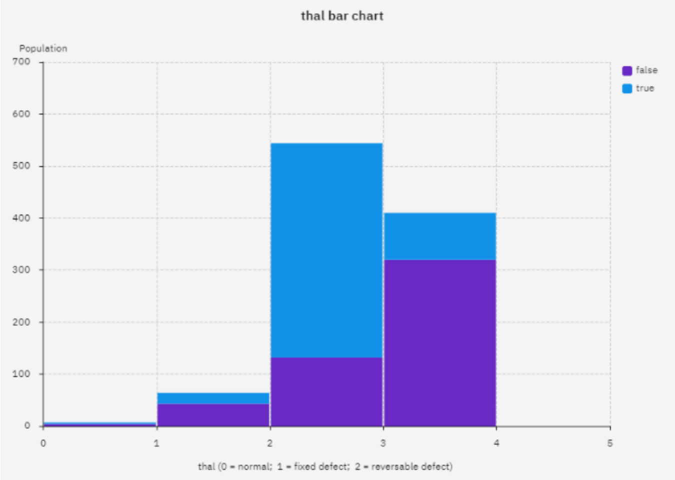


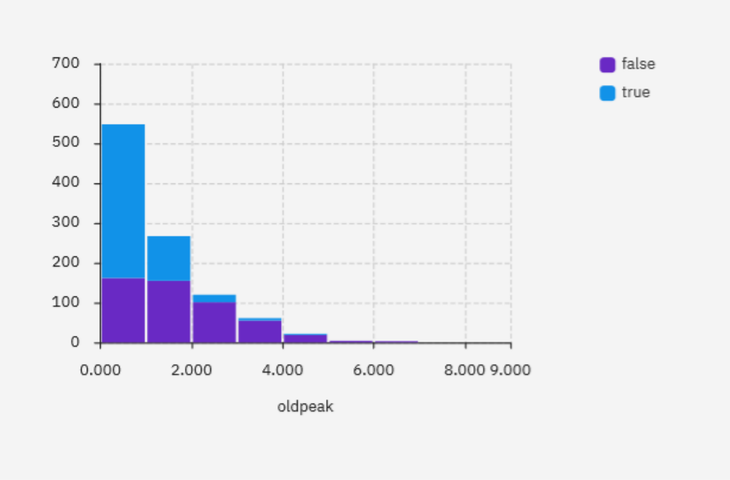


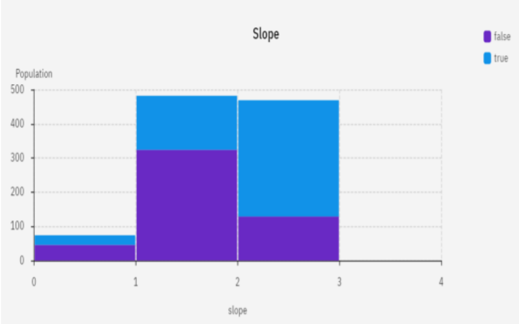
- Blood pressure, cholesterol and fasting blood sugar does not seem to have strong individual correlation with the diagnosis of CHD
- From health professionals, these are attributes which are thought to increase the likelihood of CHD
- As such we ran another feature combining all three into one.

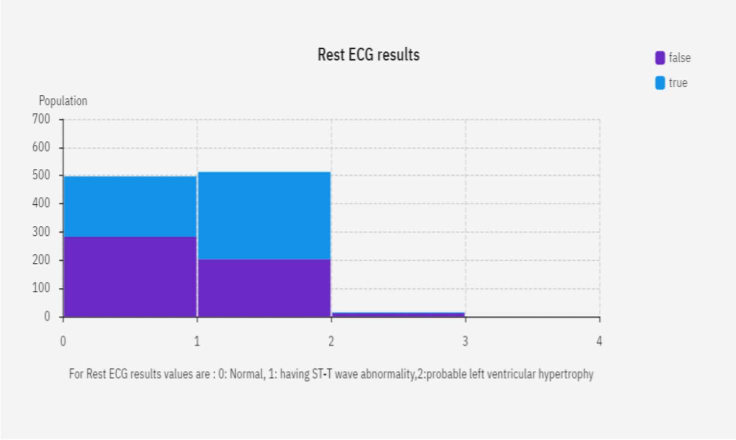
Attribute Insights and Diagnostic Analysis (H3)

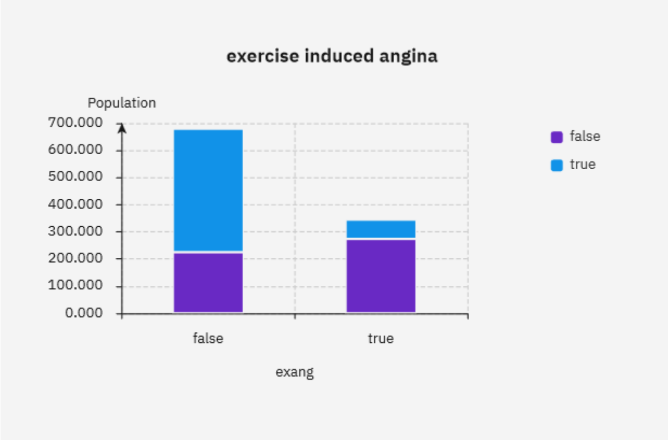


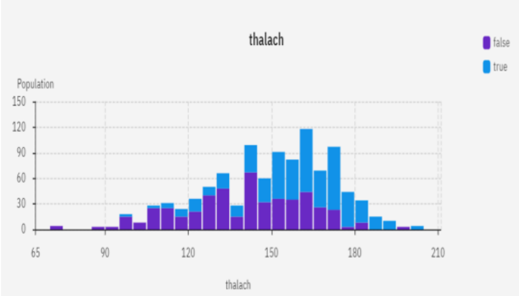












Abnormal rest ecg, higher thalach (max exercise ECG heartrate), lower oldpeak values, thalach_under_maxHealthHR is positive (ie. thalach is not under max healthy heartrate), positive slope values and no exang (no exercise induced angina) seem to indicate higher likelihood of CHD generally.

Level 2 Hypothesis (H3)

Hypothesis	Statement
H3	Patients may be diagnosed with CHD depending on the ECG-related features such as restecg, oldpeak, slope, exang, thalach and max stressed heart rates.


Predictive
Analytics

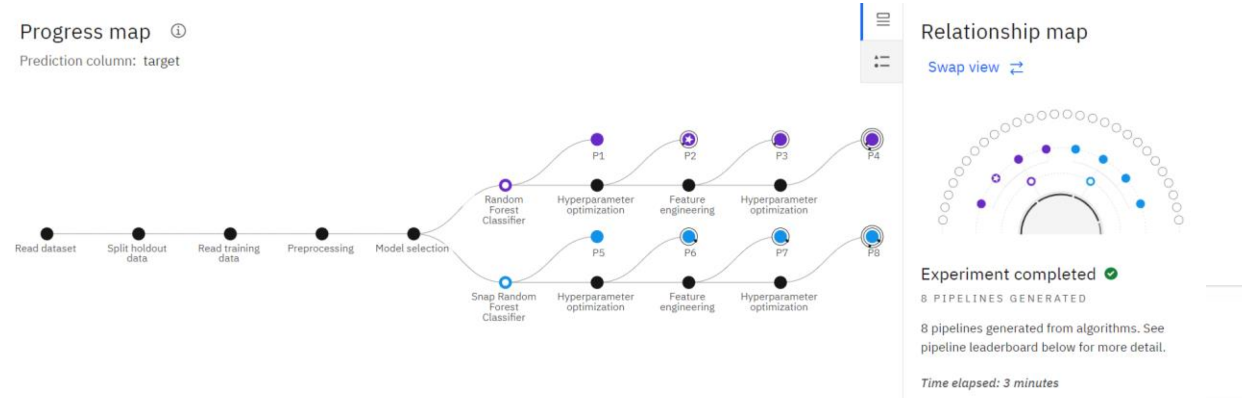
Train Data Models
(Data Modelling with Auto AI)
Hypothesis 1

The AutoAI process follows this **sequence to build the pipelines**:

- Data pre-processing
- Automated model selection
- Automated feature engineering
- Hyperparameter optimization

4 features to make the prediction:

- CP_true
- Age
- Gender
- Restecg

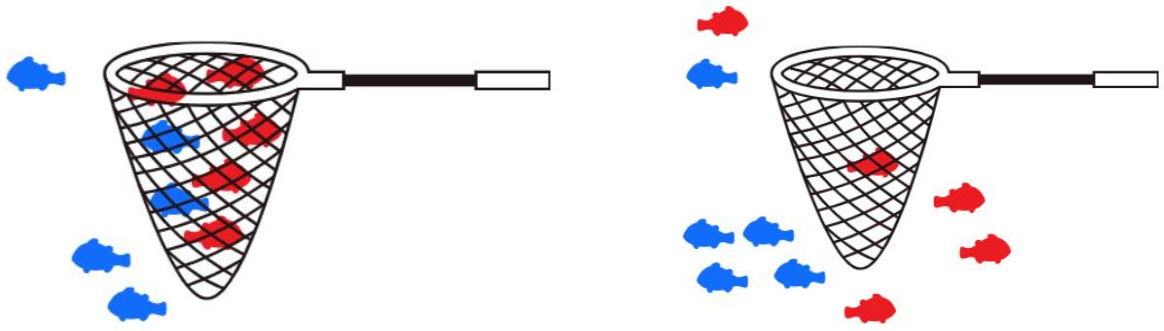


Pipeline leaderboard

	Rank	Name	Algorithm	Recall (Optimized) Cross Validation	Recall (Optimized) Holdout	Enhancements	Build time
★	1	Pipeline 2	Snap Random Forest Classifier	0.814	0.774	HPO-1	00:00:04
	2	Pipeline 3	Snap Random Forest Classifier	0.814	0.774	HPO-1 FE	00:00:25
	3	Pipeline 4	Snap Random Forest Classifier	0.814	0.774	HPO-1 FE HPO-2	00:00:10
	4	Pipeline 1	Snap Random Forest Classifier	0.797	0.811	None	00:00:01
	5	Pipeline 5	Snap Logistic Regression	0.782	0.774	None	00:00:01
	6	Pipeline 6	Snap Logistic Regression	0.782	0.774	HPO-1	00:00:01
	7	Pipeline 7	Snap Logistic Regression	0.782	0.774	HPO-1 FE	00:00:23
	8	Pipeline 8	Snap Logistic Regression	0.782	0.774	HPO-1 FE HPO-2	00:00:01

Precision vs Recall

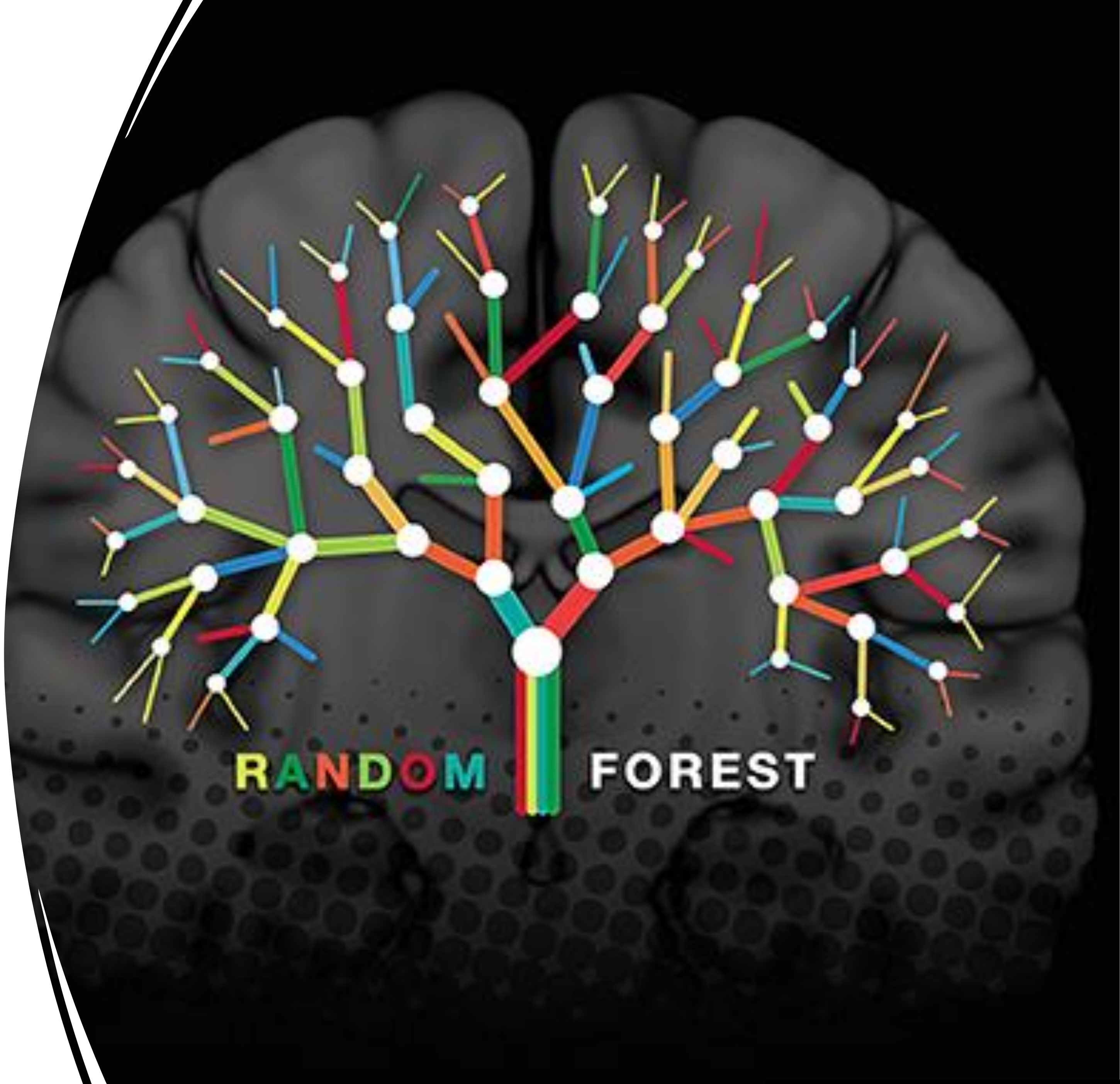
In our situation (imagine the red fish represents patient who has CHD, it is ok if we misclassify healthy patients as members of the positive class (has CHD). Missing a person who needs treatment, on the other hand, is something we don't want. As such, we want very high recall values: find as many members of the positive class as possible.



Recall **Precision**

Random Forest Classifier

It is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.



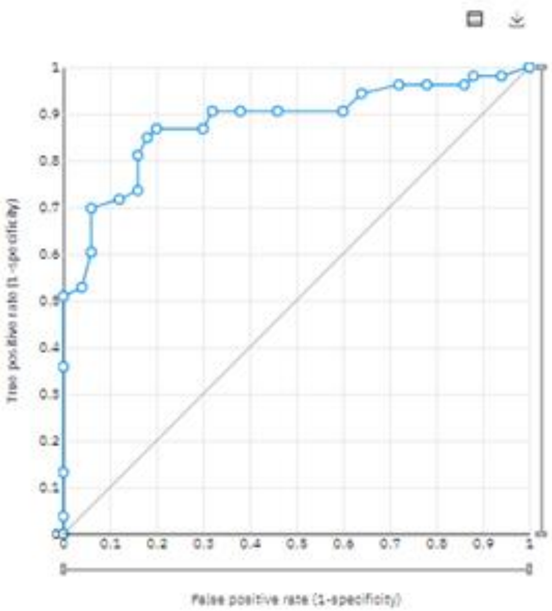
Validate Data Models – Hypothesis 1

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.806	0.787
Area under ROC	0.883	0.878
Precision	0.837	0.782
Recall	0.774	0.814
F1	0.804	0.797
Average precision	0.912	0.732
Log loss	0.486	0.477

Validate Data Models – Hypothesis 1

ROC curve



Confusion matrix

Observed	Predicted 0	Predicted 1	Percent correct
0	42	8	84.0%
1	12	41	77.4%
Percent correct	77.8%	83.7%	80.6%

Less correct — More correct

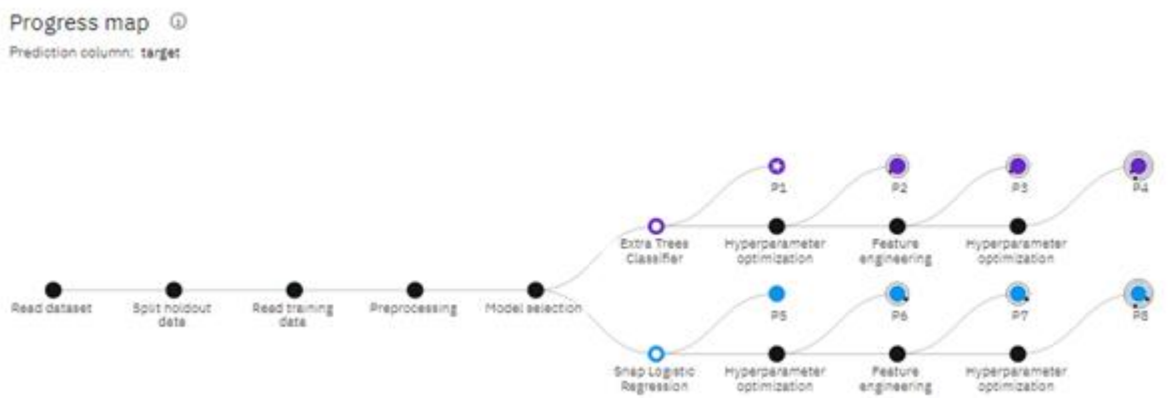
As per H1, the model can be deployed and monitored.

Improvements can also be made to the model: increase features, increase data points

Train Data Models
(Data Modelling with Auto AI)
Hypothesis 2

Single feature to make the prediction:

- chol_bps_fbs_HIGH



Pipeline leaderboard

	Rank	Name	Algorithm	Recall (Optimized) Cross Validation	Recall (Optimized) Holdout	Enhancements	Build time
★	1	Pipeline 1	Extra Trees Classifier	0.937	0.925	None	00:00:01
	2	Pipeline 2	Extra Trees Classifier	0.937	0.925	HPO-1	00:00:01
	3	Pipeline 3	Extra Trees Classifier	0.937	0.925	HPO-1 FE	00:00:07
	4	Pipeline 4	Extra Trees Classifier	0.937	0.925	HPO-1 FE HPO-2	00:00:01
	5	Pipeline 5	Snap Logistic Regression	0.937	0.925	None	00:00:01
	6	Pipeline 6	Snap Logistic Regression	0.937	0.925	HPO-1	00:00:01
	7	Pipeline 7	Snap Logistic Regression	0.937	0.925	HPO-1 FE	00:00:06
	8	Pipeline 8	Snap Logistic Regression	0.937	0.925	HPO-1 FE HPO-2	00:00:01

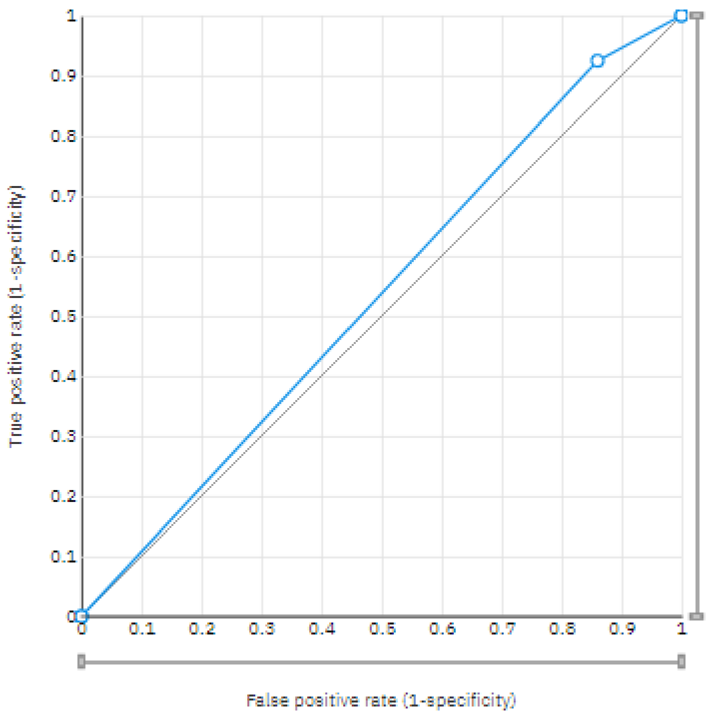
Extra Trees Classifier is also a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output it's classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest.

Validate Data Models – Hypothesis 2

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.544	0.535
Area under ROC	0.532	0.524
Precision	0.533	0.526
Recall	0.925	0.937
F1	0.676	0.674
Average precision	0.531	0.525
Log loss	0.688	0.692

Validate Data Models – Hypothesis 2



Confusion matrix

Observed	Predicted		
	0	1	Percent correct
0	7	43	14.0%
1	4	49	92.5%
Percent correct	63.6%	53.3%	54.4%

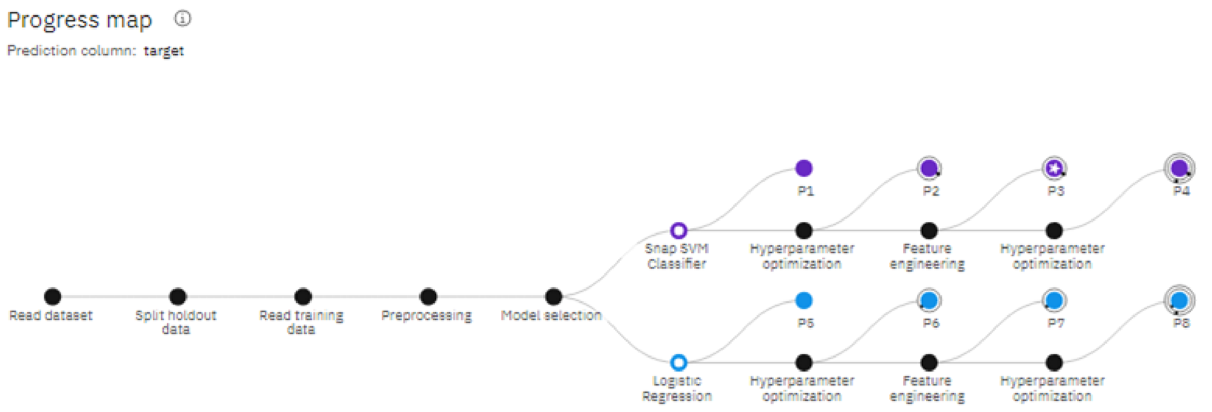
Less correct … More correct

For H2, the model is not acceptable for deployment.

H2 cannot be validated.

Train Data Models

(Data Modelling with Auto AI) - Hypothesis 3



Pipeline leaderboard

	Rank	Name	Algorithm	Recall (Optimized) Cross Validation ↓	Recall (Optimized) Holdout	Enhancements	Build time
★	1	Pipeline 3	Snap SVM Classifier	0.841	0.811	HPO-1 FE	00:00:19
	2	Pipeline 4	Snap SVM Classifier	0.841	0.811	HPO-1 FE HPO-2	00:00:01
	3	Pipeline 2	Snap SVM Classifier	0.831	0.868	HPO-1	00:00:01
	4	Pipeline 1	Snap SVM Classifier	0.831	0.868	*None*	00:00:01
	5	Pipeline 7	Logistic Regression	0.831	0.830	HPO-1 FE	00:00:19
	6	Pipeline 8	Logistic Regression	0.831	0.830	HPO-1 FE HPO-2	00:00:01
	7	Pipeline 5	Logistic Regression	0.816	0.830	*None*	00:00:01
	8	Pipeline 6	Logistic Regression	0.816	0.830	HPO-1	00:00:01

6 features to make the prediction:

- Restecg
- Exang
- Oldpeak
- Slope
- Thalach
- Thalach_under_maxhealthHR

Validate Data Models – Hypothesis 3

SVM vs Logistic Regression

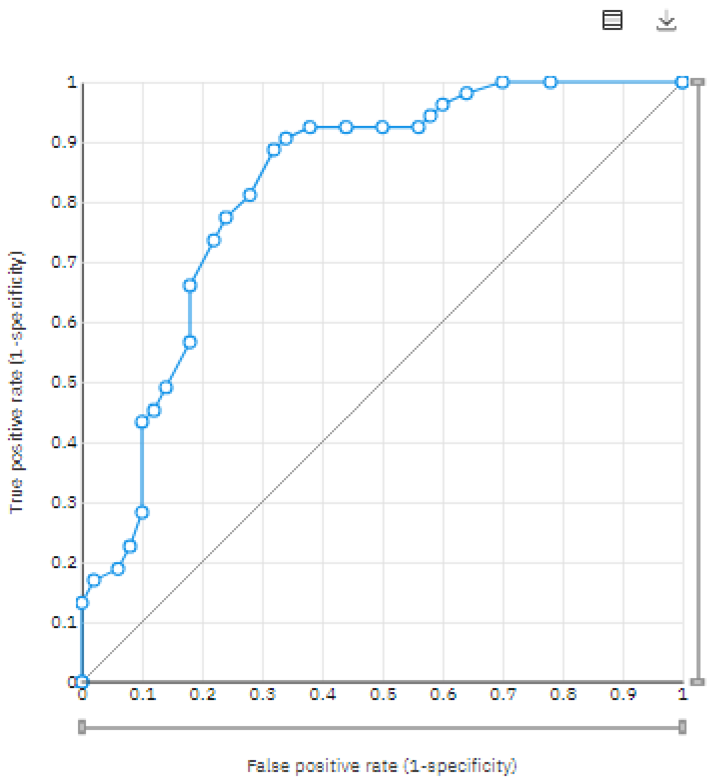
- SVM tries to finds the “best” margin (distance between the line and the support vectors) that separates the classes, and this reduces the risk of error on the data, while logistic regression does not, instead it can have different decision boundaries with different weights that are near the optimal point.
- The risk of overfitting is less in SVM, while Logistic regression is vulnerable to overfitting.

Validate Data Models – Hypothesis 3

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.777	0.794
Area under ROC	0.828	0.857
Precision	0.768	0.776
Recall	0.811	0.841
F1	0.789	0.807
Average precision	0.806	0.735

Validate Data Models – Hypothesis 3



Confusion matrix

Observed	Predicted 0	Predicted 1	Percent correct
0	37	13	74.0%
1	10	43	81.1%
Percent correct	78.7%	76.8%	77.7%

Less correct — More correct

As per H3, the model can be deployed and monitored.

Improvements can also be made to the model: increase data points

Prescriptive Analytics

Model Deployment

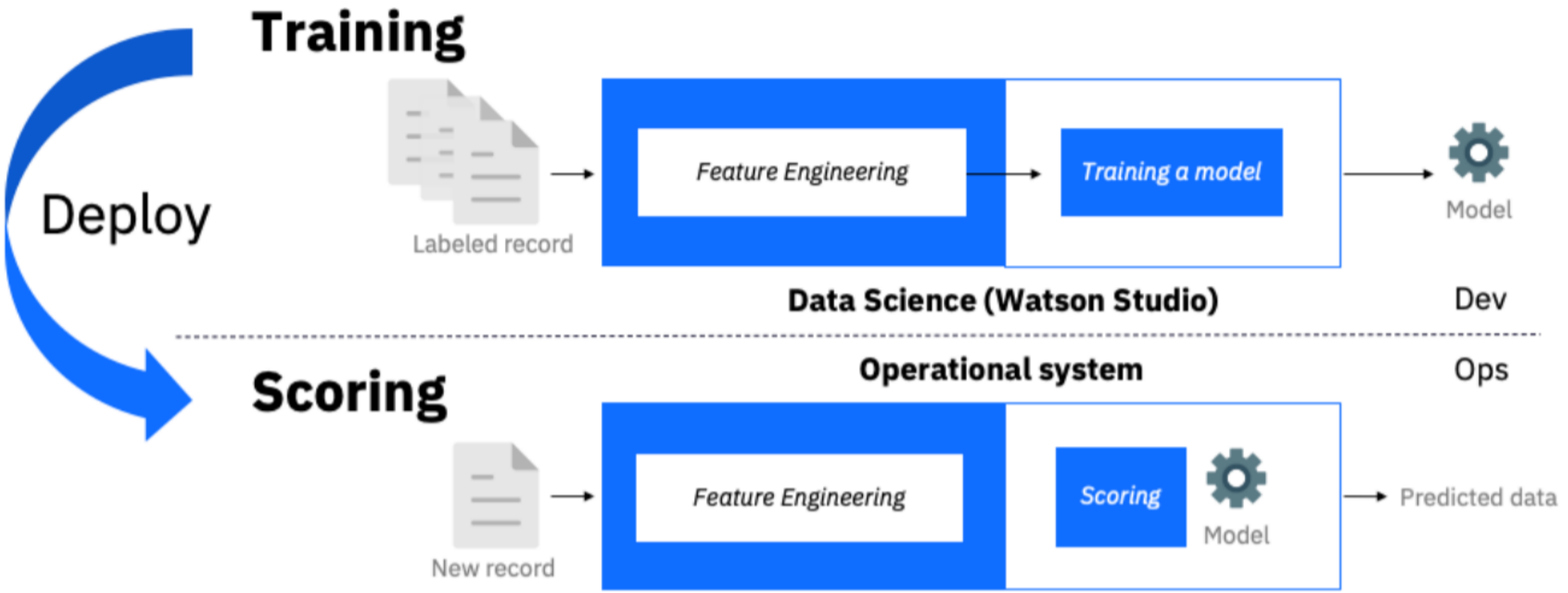
Before Deployment

The questions you should ask yourself are:

- What are my deployment requirements?
- How will I evaluate the model's performance in production?
- Have I minimized the trap of overfitting my model?
- How frequently do I plan on re-training my model?
- What are the data preprocessing needs?
- Will the format of the production input data differ drastically from the model training data?
- Will it come in batches or as a stream?
- Do I need to run the model offline?

Model Deployment

Deployment Cycle



Model Deployment

New admission / testing process to make sure patient get the result without delay

Admission with chest pain → Take Rest ECG → Monitor if patient still have chest pain → Do angiogram, fluoroscopy (which ever if preferred) to determine CHD

Model Deployment

Early Detection during Medical Check-up

Report shows individual has unusual ECG readings → Check whether the person has chest pain recently → If person has chest pain before, recommend to do angiogram / flouroscopy → Advise on costing and subsidies through Medisave

Environmental Feedback

- New data / reports will be collected from Admission Office and doctors on the rate of CHD detection after implementing the new process
- Feedback / complaints from patients related to CHD
- Additional features like family history, lifestyle, diet, etc

Summary & Reflection

Constraints and Challenges

Constraints

1. The data are collected are historical and spans more than 3 decades. It is hard to clarify on how some readings were done.
2. Data is from European and US. Comparison with other country (Asian, African, etc) is unavailable
3. Dietary change over the last 3 decades may have affected onset of CHD.
4. Family historical record was unavailable
5. 20s age group is not found in this research.
6. Small sample size for 70s age group

Constraints and Challenges

Challenges

1. Watson Studio was down for 2 working days as such, analysis was delayed. Manual computations and visualization with Jupyter Notebook using Python was used during the outage.
2. A very steep learning curve for some of the members in understanding data transformation and models to deploy in prescriptive.
3. Additional time spent in exploring Auto AI and different models to deploy
4. Sometimes Auto AI choose models that shows over high scoring. As such we need to reset the preferred models for running our hypotheses
5. Public understanding and awareness of CHD are from online resources / hearsay.
6. Healthcare professionals' support to this change initiative.
7. Getting staff to attend for training and getting the process right.

Conclusion

- Predicting CHD is a challenging and ongoing.
- The advancement of medical science and data science can help to make assessment and treatment more readily and accurately.
- Understanding the symptoms that most likely related to CHD will help healthcare professions quickly diagnose the disease and ultimate save the lives of people who are affected by it.
- Collection of more data would increase confidence in predictive performance
- Have a deep understanding of certain features and how they might interact.
- Close working with SME will help identify important features

Future Enhancement

- Getting Public Health Agencies to support in CHD aware / brochures
- Finding related features not recorded in the research to better determine CHD
- More support from government subsidies in more expensive testing like angiogram
- Collaboration with other teams than may have done similar research and share data and knowledge

Key Take-away from the Project

- Although using Watson Auto AI save time from doing the manual way through Jupyter Notebook, settings must be selected carefully in order to run the pipeline correctly.
- Be open-minded, supportive and learn from one another.
- Have a better understanding of the challenges faced at each stage of the DS Lifecycle, during the Project.
- Appreciate the dynamics and support of the diverse Team in contributing and sharing their knowledge and skills.