Bronco ID: 016065686
Last Name: Liang
First Name: Johnny

Question 1

Similar to the description, the model shown in the diagram has a measure of performance (P) of task (T) as represented by the 'evaluate' block. While the 'analyze errors' block implies that it learns from experience, it does not necessarily improve based on its performance, because the model goes to 'launch' when it has a positive performance.

Question 2

Preprocessing - Preparing data so that only complete, accurate and useful data is used

Machine learning - This is where the knowledge building off of that data happens

Postprocessing - Making use of the previous steps by visualizing or interpreting results

Question 3

a. **Data distribution** - The test data's distribution has less deviation compared to the training data, which challenges its ability to evaluate the performance.
b. **Outlier** - There's a datapoint that exists far from the data trend
c. **Missing values** - Data has a lot of missing / non-numeric values
d. **Outliers** or **Noise** - Data has an area where the data does not follow the typical trend, either due to being outliers or noise making the data inaccurate.
e. **Sparsity** - Data contains a lot of zeros, with an entire column and row being all 0

Question 4

a. To record and analyze people's eyesight data and its relation to being recommended contact lens
b. A feature is a data column that describes a characteristic, for example age
c. A feature value is a specific value of an instance's feature, for example, the first instance has an age value of 'young'
d. Dimensionality is the number of features that exist in a dataset. Our dataset has 5
e. An instance is a row. It represents a person or object that the features apply to.
   For example, our dataset's first instance is young, myope, has no astigmatism, reduced tear production and is not recommended for contact lens.
f. A class is a grouping of instances that all have a certain feature,
   for example, in our data people are either classified Myope or Hypermetrope

Question 5

a. Supervised learning. All the data is labeled (X / O).
b. Semi-supervised learning. Some data are labeled (X / O), while most are unlabeled.
c. Unsupervised learning. All data is unlabeled.

Question 6

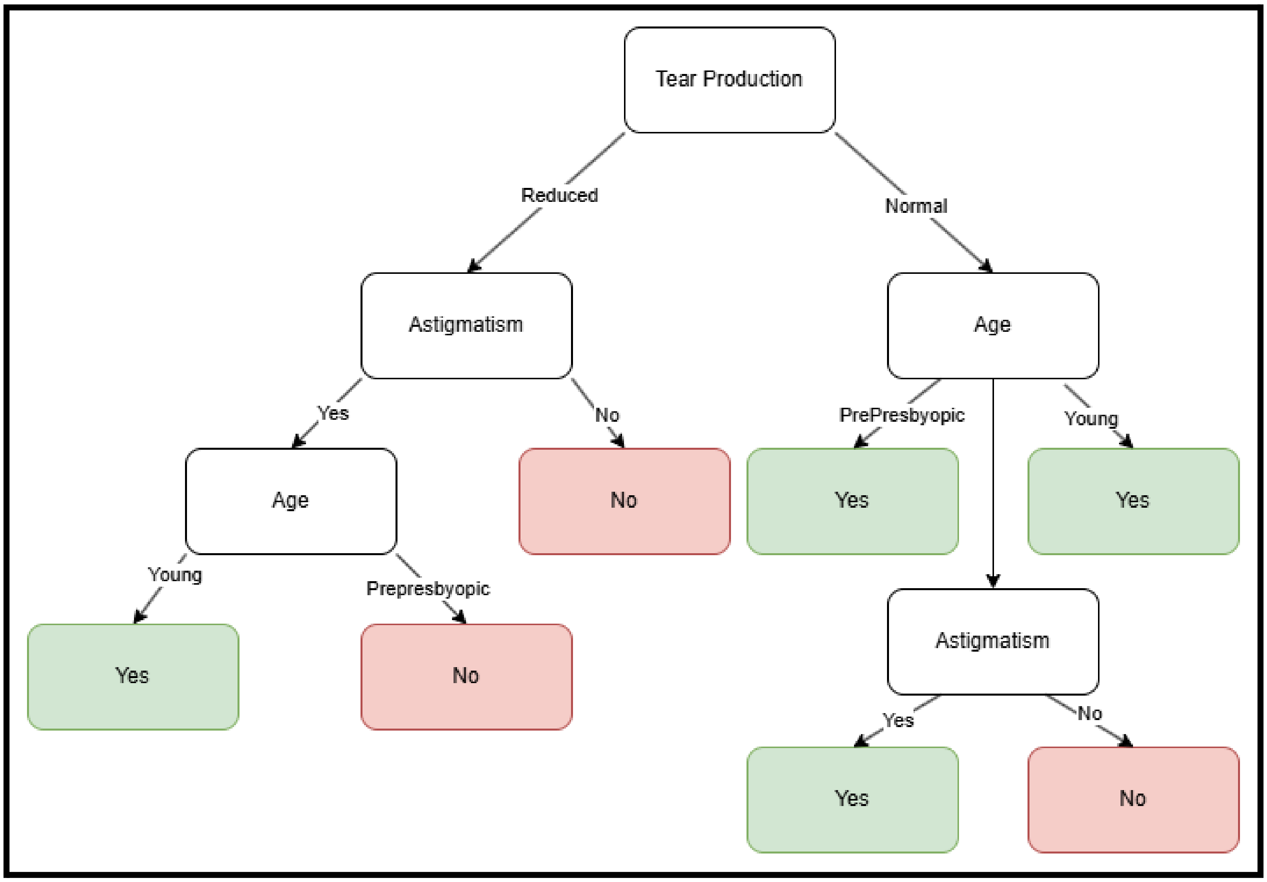
Binary classifier - For tasks where K is a boolean, w/ each concept being in 1 of the classes C

Multi-class classifier - For tasks where the dataset has multiple classes K, and each object has one class C.

Multi-label classifier - For tasks where the dataset has multiple classes K, and each object in the dataset can have multiple classes C at once.

Question 7

a. Final decision tree. ***Work below**.



b. https://github.com/cheeseb0rger/CS4210_Assignment_1

c. The only significant difference is that the program only created branches of 2s. Meanwhile, my ‘Age’ node has 3 branches since Age has 3 features. There were also multiple instances where the information gain was the same for 2 values, which could also influence the results.

---

entropy(S) = – 4/10 log2 4/10 – 6/10 log2 6/10 **= 0.9709**

gain(S, age) = 0.9709 - (4/10)entropy(S_{young}) - (3/10)entropy($S_{preby..}$) - (3/10)entropy($S_{prepreby..}$)
gain(S, age) = 0.9709 - 0.4 - 0.2755 - 0.2755 **= 0.0199**

gain(S, spectacle) = 0.9709 - (8/10)entropy(S_{Myope}) - (2/10)entropy($S_{Hypermetro}$)
gain(S, spectacle) = 0.9709 - 0.8 - 0 **= 0.1709**

gain(S, astigmatism) = 0.9709 - (4/10)entropy(S_{Yes}) - (6/10)entropy(S_{No})
gain(S, astigmatism) = 0.9709 - 0.3245 - 0.3900 **= 0.2564** ←

gain(S, tear) = 0.9709 - (6/10)entropy($S_{Reduced}$) - (4/10)entropy(S_{Normal})
gain(S, tear) = 0.9709 - 0.3245 - 0.3900 **= 0.2564** ←

*Both gain(S, astigmatism) and gain(S, tear) has the highest information gain…
I will choose tear production as the root.

$S_{Reduced}$ = [+1, -5]
entropy($S_{Reduced}$) = 0.6500

gain($S_{Reduced}$, age) = 0.6500 - (3/6) 0.9183 - (1/6) 0.0 - (2/6) 0.0 **= 0.1377**
gain($S_{Reduced}$, spectacle) = 0.6500 - (2/6) 0.0 - (4/6) 0.8112 **= 0.1092**
gain($S_{Reduced}$, Astigmatism) = 0.6500 - (4/6) 0.0 - (2/6) 1 **= 0.3166** ←

Astigmatism will branch from Tear production - Reduced

S_{Normal} = [+3,-1]
entropy(S_{Normal}) = 0.8112

gain(S_{Normal}, age) = 0.8112 - (1/4) 0.0 - (2/4) 1.0 - (1/4) 0.0 **= 0.3112** ←
gain(S_{Normal}, spectacle) = 0.8112 - (4/4) 0.8112 **= 0.0**
gain(S_{Normal}, Astigmatism) = 0.8112 - (2/4) 1.0 - (2/4) 0.0 **= 0.3112** ←

* again, we have 2 of the same values for information gain
I will choose age to branch from Tear production - Normal

$S_{Reduced\ \&\ No\ Astigmatism}$ = [+0,-4] → No
No (leaf) branches from Astigmatism - No

$S_{Reduced\ \&\ Yes\ Astigmatism}$ = [+1,-1]
entropy($S_{Normal\ \&\ No\ Astigmatism}$**) = 1.0**

gain($S_{Normal\ \&\ No\ Astigmatism}$**, age)** = 1.0 - (1/2) 0.0 - (1/2) 0.0 = 1.0
gain($S_{Normal\ \&\ No\ Astigmatism}$**, spectacle)** = 1.0 - (2/2) 0.0 = 1.0

* again, we have 2 of the same values for information gain

I will choose <u>Age</u> to branch from <u>Astigmatism - Yes</u>

Which results in:

Young→ Yes

Prepresbyoptic → No * there are no instances left for age = Presbyopic..

$S_{\text{Normal \& Young}}$ = [-0,+1] → Yes

Yes (leaf) branches from Age - Young

$S_{\text{Normal \& Prepresbyoptic}}$ = [0,+1] → Yes

Yes (leaf) branches from Age - Prepresbyopic

$S_{\text{Normal \& Presbyoptic}}$ = [-1,+1]

entropy($S_{\text{Normal \& Presbyoptic}}$) = 1.0

Gain($S_{\text{Normal \& Presbyoptic,}}$ Age) = 1.0

Gain($S_{\text{Normal \& Presbyoptic}}$ Spectacle) = 1.0

* again, we have 2 of the same values for information gain

I will choose <u>Astigmatism</u> to branch from <u>Age - Presbyoptic</u>

Which results in:

Yes Astigmatism → Yes

No Astigmatism → No