1.

![Graphical user interface, text, application, email

Description automatically generated]()

2.

![Graphical user interface, text, application, email

Description automatically generated]()

3. Looking at the performance of the model on the training and test sets and make an argument on whether the model is overfitting the training set.

Based on the data, I would say that the model is not overfitting the training set because the accuracy of both are so close. We know that overfitting can look like near accuracy on the training set and have lower performance on the test set. Because the training set doesn’t have near perfect accuracy and the two sets in the model have a relatively close accuracy, it is less likely that there is overfitting.

4. Look at the performance of the models and state which model you prefer. Propose an explanation of why one feature was better than the other.

I prefer the decision tree model, I like having the option to add more classifications. The feature for male was better than the others because it is easier to split into a binary classifier. For height and weight, you would have to pick sections to place the heights and weights in, then turn it into a binary classifier.

5. Which features would you propose are the best to train on? Something to explore later: Does age improve the predictive power of our decision tree?

I believe the best features to train on are the ones that are easier to make binary classifiers out of. I would say in this set it would be gender and age. I do believe that age would improve our predictive power because of the variable that weight is affected by age and visa versa.

6.

![Chart, scatter chart

Description automatically generated]()

7.

Chart

Description automatically generated

8.

![Graphical user interface, text, application, email

Description automatically generated]()

9. How does the SVC model (using RBF by default) compare with the other models?

This model depends on the hyperplane and where each instance falls around it. If an instance is closest to the hyperplane or on the wrong side of it, it means that we are less confident about their classification. This model’s performance also depends on the number of instances within the set. This model is also susceptible to overfitting.

10.

Chart, scatter chart

Description automatically generated

• Where are the boundaries of the positive/negative regions?

The top right and the bottom left of the plot.

• Where is the region that predictions should be considered tentative?

The center of the plot as they are so close and can be susceptible to inaccurate predictions or overfitting.

11.

![Graphical user interface, text, application, email

Description automatically generated]()

12. How does the NN model compare with the other models?

This model depends on the input and output layers being fully connected. This model uses weights wherein we feed in the number of instances and measure the difference between desired and computed values. The weights can be adjusted, and the learning rate is determined by this. Overfitting is an issue with this model as well and can negatively affect its interpretability. This model has limited performance. There can be thousands of internal parameters within this model, which can cause the training to be slow and inaccurate.