Submission 1: Preparation Code

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, scatter chart

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart, scatter chart

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Submission 2: Training and Test Performance with regular split

Graphical user interface, text, application

Description automatically generated

Submission 3: Analysis of Overfitting

While the accuracy and f-score dropped a small bit from the training results to the test results, this drop was not very significant. I would say that overfitting is not a big issue when it comes to this data and using height as an input. Note: I used the regular split.

Submission 4: Performance of new model

1. The accuracy and f-score for the training set with weight as the input is very high, but when we get to the test set there is a large decrease in both accuracy and f-score. There may be some overfitting or some other issue going on here. I would prefer to use heights for the model that we will use to make our predictions. It was much more consistent in results from training to test sets.
2. I would say that for the most part men are taller on average than women. While, overall men probably weigh more than women, there is much more fluctuation in weights among men and women, with some men being very tall and skinny to some women being short and weighing more.

Submission 5: Analysis of which feature to train on

Just looking at the different results, I would probably use height to train on. I feel as though the consistency of high accuracy and f-scores between training and test scores is a good indicator that this would be a good feature to use.

Submission 6: Scatterplot

Chart, scatter chart

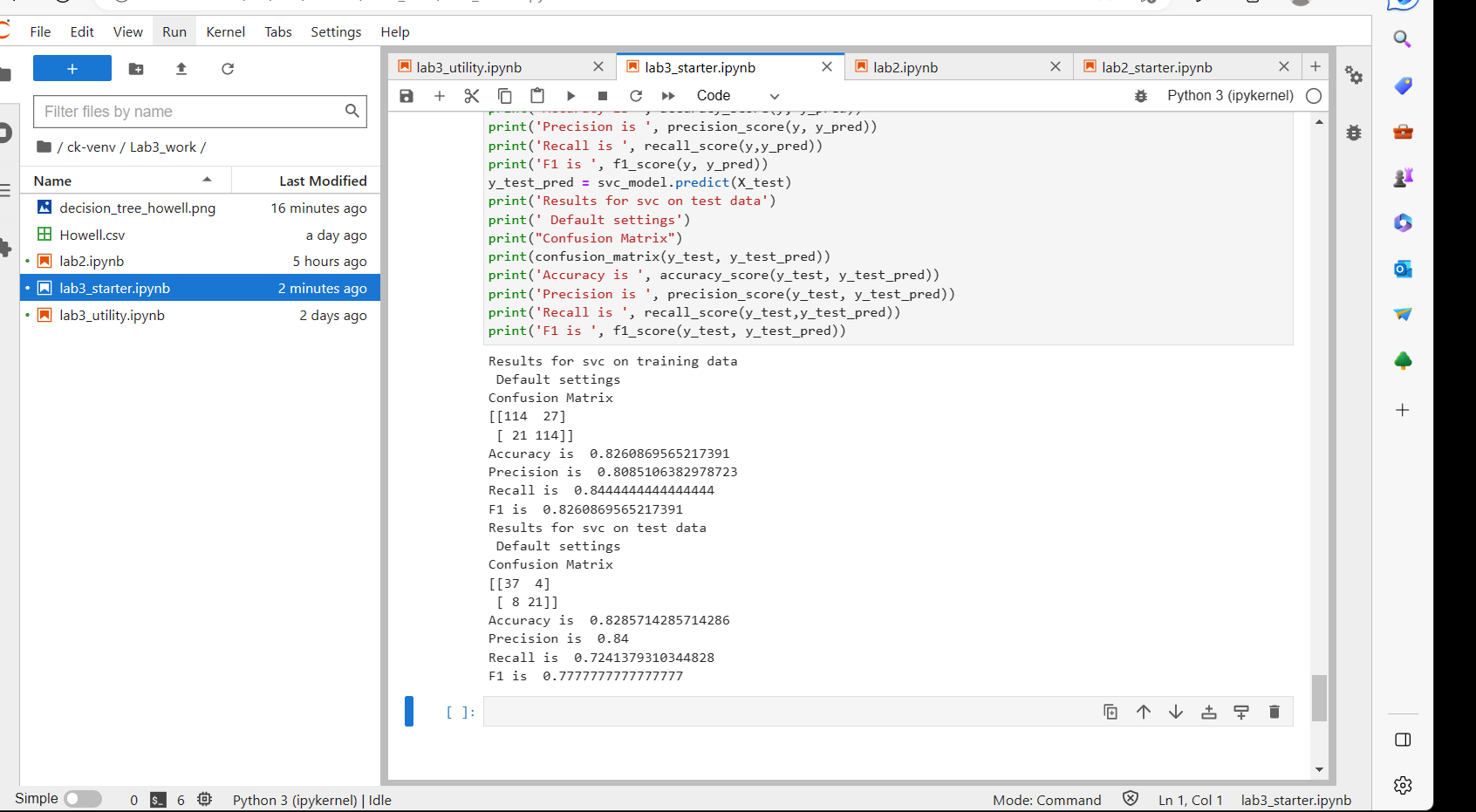
Description automatically generated

Submission 7: Tree Plot

Chart, scatter chart

Description automatically generated

Submission 8: Performance for SVC



Submission 9: Comparison

The accuracy and f-scores for the SVC model were lower than the decision tree when height was the input. When weight was the input for the decision tree the SVC model had a lower accuracy and f-score for the training set, but it was higher on the test set. When taking into account the height and weight, the SVC model was definitely lower on the training set for accuracy and f-score, but pretty comparable when it came to the test set.

Submission 10: Scatter with Support Vectors

Graphical user interface, chart, scatter chart

Description automatically generated

1. It looks like the positive/negative boundary for the females is where x = 145 and the positive/negative value for the males is around x = 161 or x = 162.
2. I would say the region around x = 155 where there is a mixed of females and males. These x-values are height in centimeters.

Submission 11: Neural Net Performance

Graphical user interface, text, application, email

Description automatically generated

Submission 12: Comparison

The neural net performance was pretty consistent between the training set and the test set. The neural network was comparable to the other models. The only other model that the neural network scored lower than with accuracy and f-scores across the board was the decision tree model using height. It had a little bit better accuracy and f-scores than the SVC. It’s performance was lower on the training sets than the decision tree using weight and using height and weight, but it outperformed these models when it came to the test set.