Question 1 “Decision Tree”

For training set of size 25% (109 row): it depends on the Random Sample that function “dataset. Sample(frac=0.25)” generate

- Minimum accuracy achieved during 5 time of runs $\cong 84\%$
- Maximum accuracy achieved during 5 times of runs $\cong 97\%$

And to tree size

- Minimum Tree Size achieved during 5 time of runs = 3 level
- Maximum Tree Size achieved during 5 time of runs = 15 levels

For training set of size 30%: we use random seeds (100-170-225-270-300) to get same result and get over variance of random samples

- Accuracies = 89.18%, 96.06%, 96.0655%, 86.229%, 95.73%
- Minimum Accuracy = 86.229%
- Max Accuracy = 96.0655%
- Average Accuracy = 92.6557%
- Tree Sizes = 11, 3,7,15,5
- Min Tree Size 3
- Max Tree Size 15
- Mean Tree Size 8.2

For training set of size 40%: we use random seeds (100-170-225-270-300) to get same result and get over variance of random samples

- Accuracies = 93.87%, 96.17%, 96.55%, 87.74%, 97.32%
- Minimum Accuracy = 87.74%
- Max Accuracy = 97.32%
- Average Accuracy = 94.329%
- Tree Sizes = 7,7,7,15,9
- Min Tree Size 7.0
- Max Tree Size 15.0
- Mean Tree Size 9.0

For training set of size 50%: we use random seeds (100-170-225-270-300) to get same result and get over variance of random samples

- Accuracies = 95.392%, 97.235%, 96.313%, 98.156%, 97.6958%
- Min Accuracy = 95.392%
- Max Accuracy = 98.156%
- Mean Accuracy = 96.958%
- Tree Sizes = 9,9,9,11,9
- Min Tree Size 9.0
- Max Tree Size 11.0
- Mean Tree Size 9.4

For training set of size 60%: we use random seeds (100-170-225-270-300) to get same result and get over variance of random samples

- Accuracies = 95.977%, 97.126%, 96.551%, 98.275%, 97.1264%
- Min Accuracy = 95.977%
- Max Accuracy = 98.275%
- Mean Accuracy = 97.0115%
- Tree Sizes = 11,11,11,11,9
- Min Tree Size 9.0
- Max Tree Size 11.0
- Mean Tree Size 10.6

For training set of size 70%: we use random seeds (100-170-225-270-300) to get same result and get over variance of random samples

- Accuracies = 96.946%, 96.183%, 96.946%, 99.236%,98.4732%
- Min Accuracy = 96.183%
- Max Accuracy = 99.236%
- Mean Accuracy = 97.56%
- Tree Sizes = 11,11,13,13,11
- Min Tree Size 11
- Max Tree Size 13
- Mean Tree Size 11.8

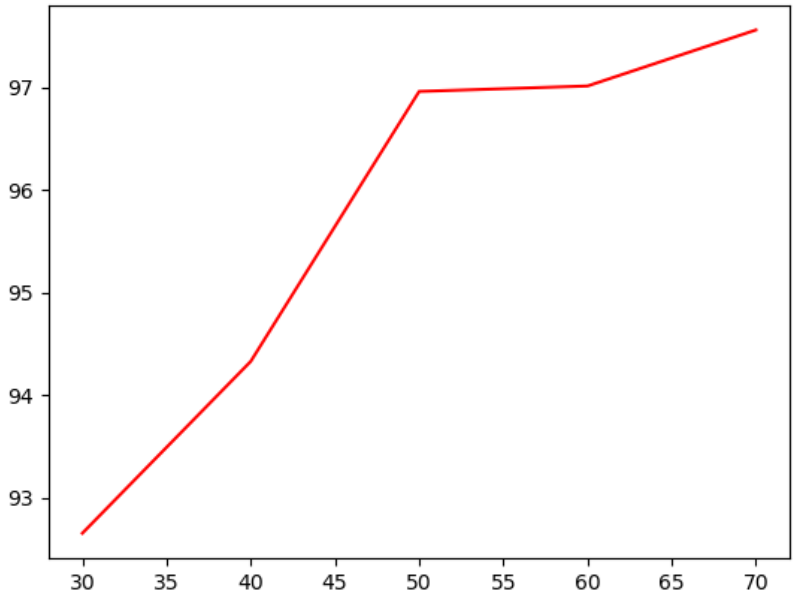

Figure 1 Train Set Size and Mean Accuracy

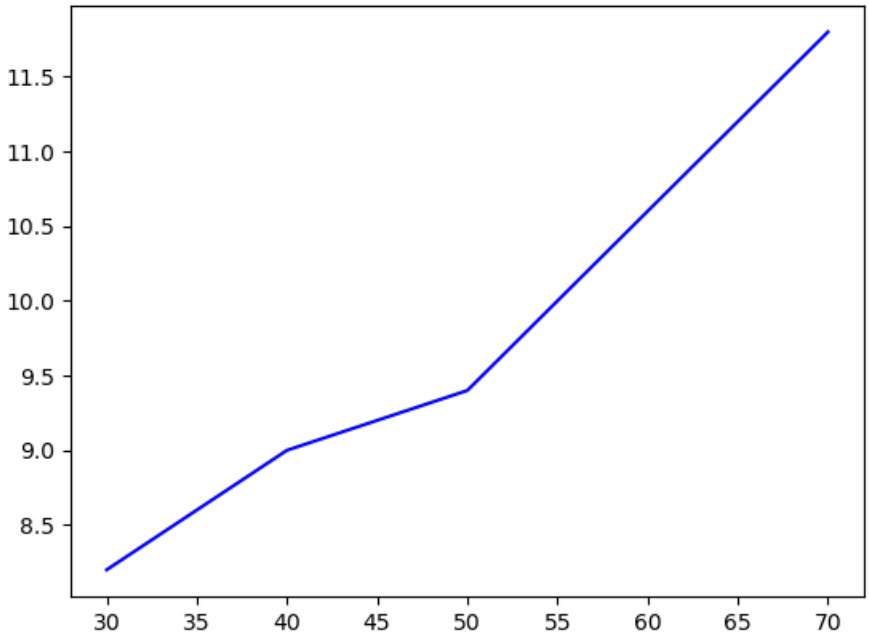

Figure 2 Train Set Size and Mean of Tree Size

Question 2 “SVM”

- We tried different set of features and choosing the best features.
- The best features= [age, chol, thalach, trestbps]
- When we drop this set [ca,restecg]the accuracy increased.
- We tried different learning rate we found that the best learning rate is 0.0001, when we are increasing the rate for example to 0.01 and decreasing it to 0.000001 the accuracy decreased.
- The accuracy of SVM = 94.67%