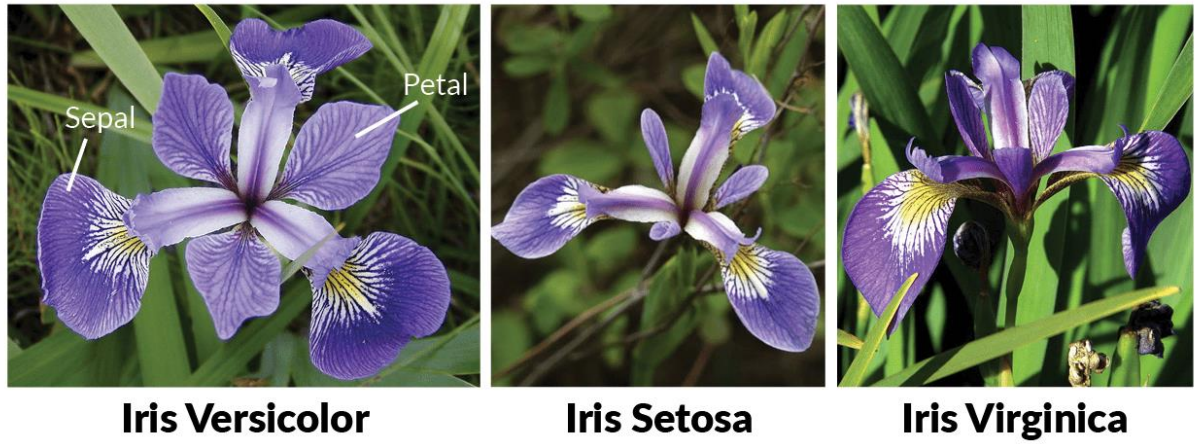
Application of Data Mining Techniques

Amare Gared

Iris Dataset

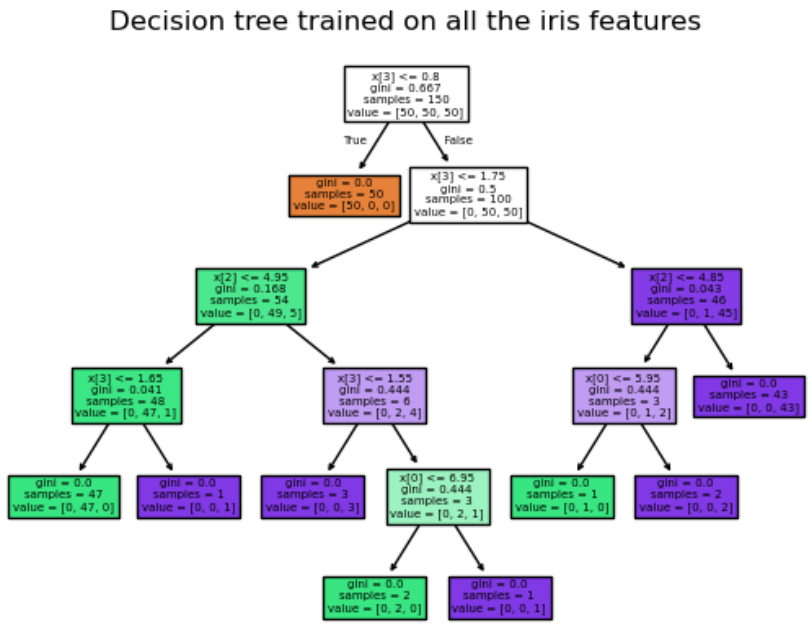
- 150 instances
- 4 Features
  - Sepal Length
  - Sepal Width
  - Petal Length
  - Petal Width

- Challenges?

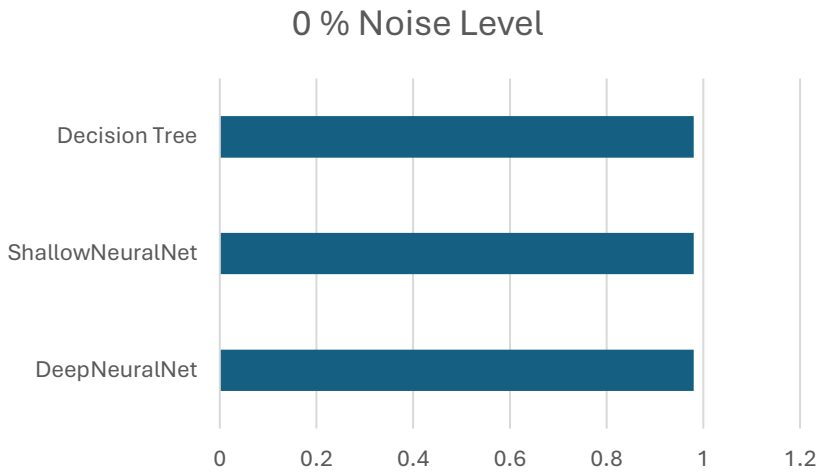


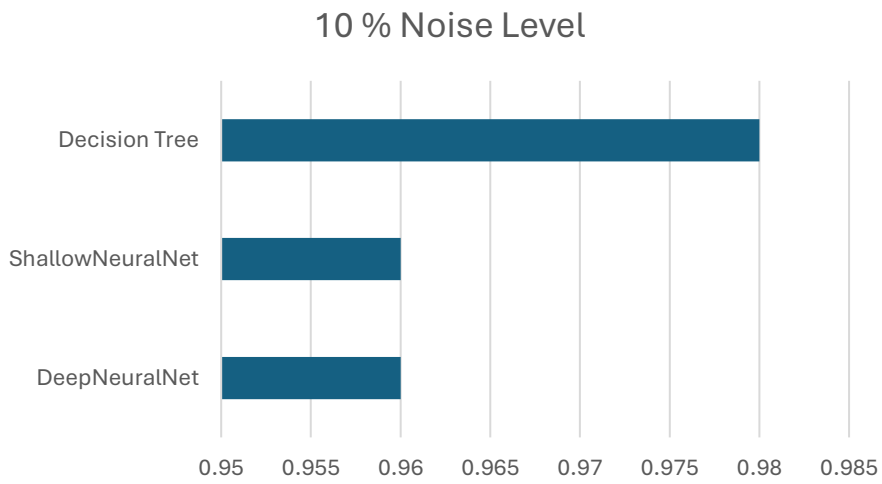
Models

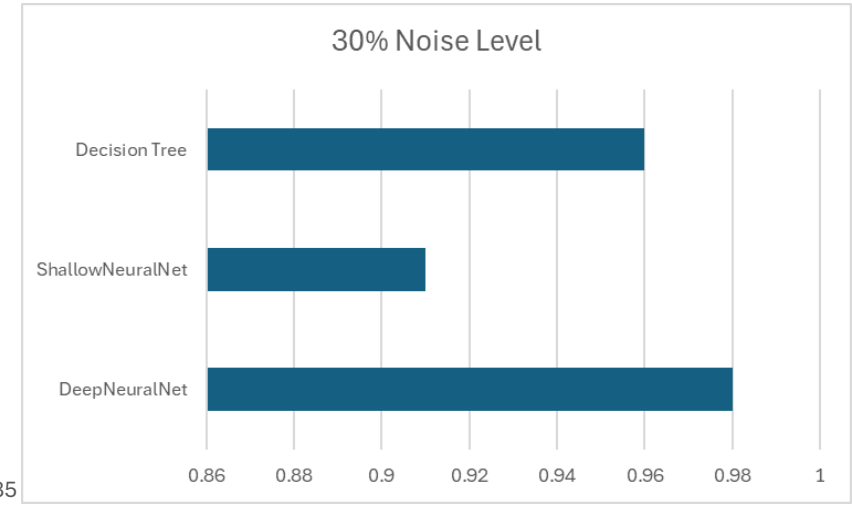
- Decision Tree
- Neural Network
  - Shallow NN
    - 1 hidden layer
  - Deep NN
    - 3 Hidden layers
- Added Noise



Results


0 % Noise Level
Decision Tree
ShallowNeuralNet
DeepNeuralNet
0
0.2
0.4
0.6
0.8
1
1.2


10 % Noise Level
Decision Tree
ShallowNeuralNet
DeepNeuralNet
0.95
0.955
0.96
0.965
0.97
0.975
0.98
0.985


30% Noise Level
Decision Tree
ShallowNeuralNet
DeepNeuralNet
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1

Conclusion

- **Impact of Noise**
  - Adding noise highlighted the differing strengths of the models. The Decision Tree excelled at maintaining accuracy with smaller, clean datasets, while the Deep Neural Network outperformed when complexity increased.
- **Performance**
  - Both the Shallow and Deep Neural Networks showed strong performance on the clean dataset but displayed varying degrees of resilience to noise. The Deep Neural Network maintained a higher accuracy at 30% noise compared to the Shallow Neural Network, indicating that its additional layers enable better generalization in noisy environments.
- **Robustness :**
  - The Decision Tree consistently demonstrated high accuracy, even at 30% noise, suggesting that it is less sensitive to noisy training data and it can be highly effective for well-structured clean datasets.