Decision Tree
A comparison & Conclusion

இப்படத்தில் வரும் சம்பவங்கள் யாவும் கற்பனையே..
நிஜத்தின் அளவிற்கு கொடூரமானவை அல்ல..

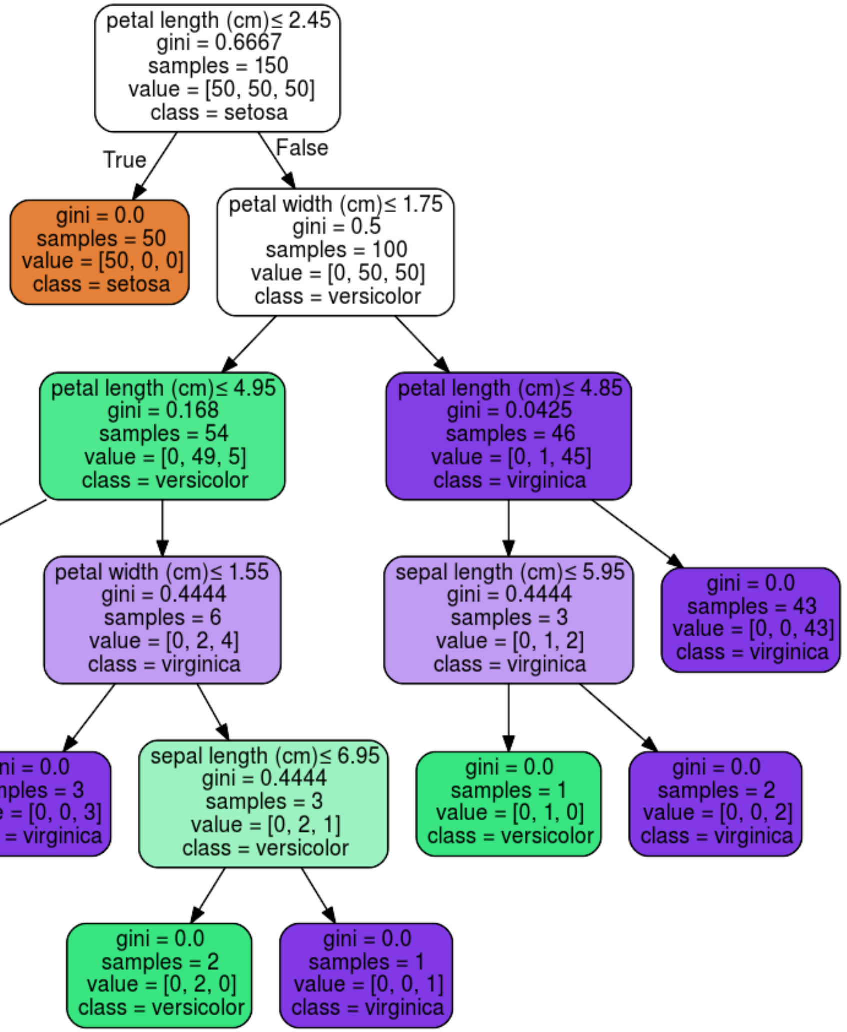
இதுவரை – ID3 vs C4.5 vs CART

Decision Trees Comparison

Features	*ID3*	*C4.5*	*CART*
Type of data	Categorical	Continuous and Categorical	continuous and nominal attributes data
Speed	Low	Faster than ID3	Average
Boosting	Not supported	Not supported	Supported
Pruning	No	Pre-pruning	Post pruning
Missing Values	Can't deal with	Can't deal with	Can deal with
Formula	Use information entropy and information Gain	Use split info and gain ratio	Use Gini diversity index

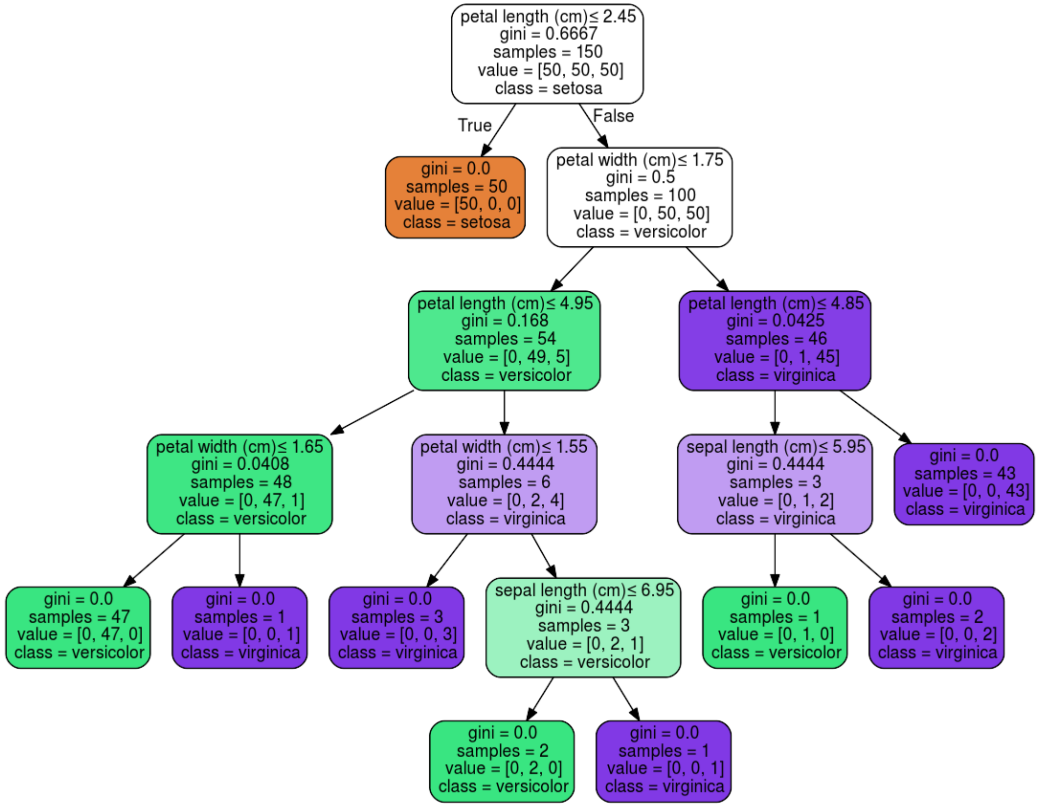
Advantages of Decision Trees

- Simple to understand, interpret, visualize.



Advantages of Decision Trees

- Decision trees *implicitly perform variable screening* or *feature selection*.



https://www.youtube.com/watch?v=wpNl-JwwplA

Advantages of Decision Trees

- Can *handle both numerical and categorical data*. Can also *handle multi-output problems.*

Advantages of Decision Trees

- Decision trees require relatively *little effort from users for data preparation*.

Advantages of Decision Trees

- People tend to *believe* Decision Trees because we can give *justification* on why model choose to give an output

Disadvantages of Decision Trees

- Decision-tree learners *can create over-complex trees* that do not generalize the data well. This is called *overfitting*.

Disadvantages of Decision Trees

- Decision trees can be unstable because *small variations in the data might result in a completely different tree being generated.* This is called ***variance***, which needs to be *lowered by methods like **bagging** and **boosting***.

Disadvantages of Decision Trees

- Greedy algorithms cannot guarantee to return the globally optimal decision tree. This can be mitigated by training multiple trees, where the features and samples are randomly sampled with replacement.

Disadvantages of Decision Trees

- Decision tree learners create *biased trees if some classes dominate*. It is therefore recommended to balance the data set prior to fitting with the decision tree.

Decision Tree Applications

- Business Management
- Customer Relationship Management
- Fraudulent Statement Detection
- Energy Consumption
- Fault Diagnosis
- Healthcare Management