Decision Trees

10 March 2024 22:06

Why decision trees ?

1. Non Linearity:

Unlike regression models, which assume a linear relationship between the predictors and the target, decision trees can handle complex interactions and non-linearities.

It can do without needing for any transformation of the features.

Decision trees can capture non-linear relationships between predictors and the target.

2. Interpretability :

Decision rules represented by the tree structure are easy to understand and visualize, making them accessible to executives who are non-experts.

[VP, Director] ← logit log(odds)

DTs are like a flow chart and business users love flow chart.

3. Scalability:

Decision trees are computationally efficient and scalable to large datasets, making them suitable for real-time applications and large-scale data processing

Training dataset = 50M Rows → on big data

~~Linear / logistic~~

↓

DT

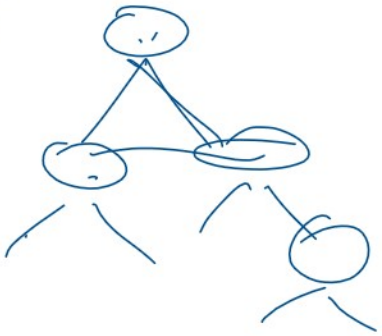
4. Handling Mixed Data Types

DTs can handle both numerical and categorical data without the need for one-hot encoding

5. Robustness to outliers and irrelevant features

DTs can handle outliers & noise in the data along with irrelevant features without significantly impacting the model performance.

[Gini impurity, Entropy]
↳ to evaluate the important features



6. Automatic selection of features

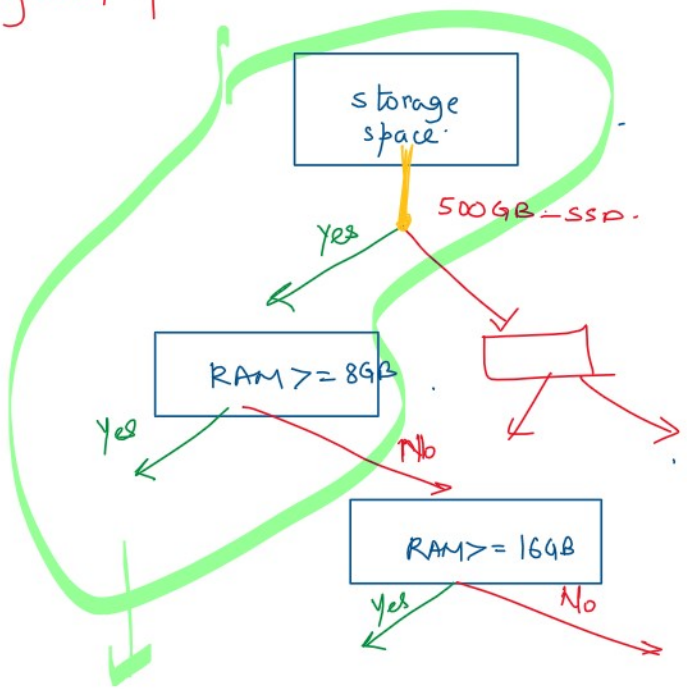
DTs perform automatic feature selection by identifying the most important features at each split.

Intuition behind Decision Trees

"Decision trees are everywhere".

Senthil → is trying to buy a laptop.

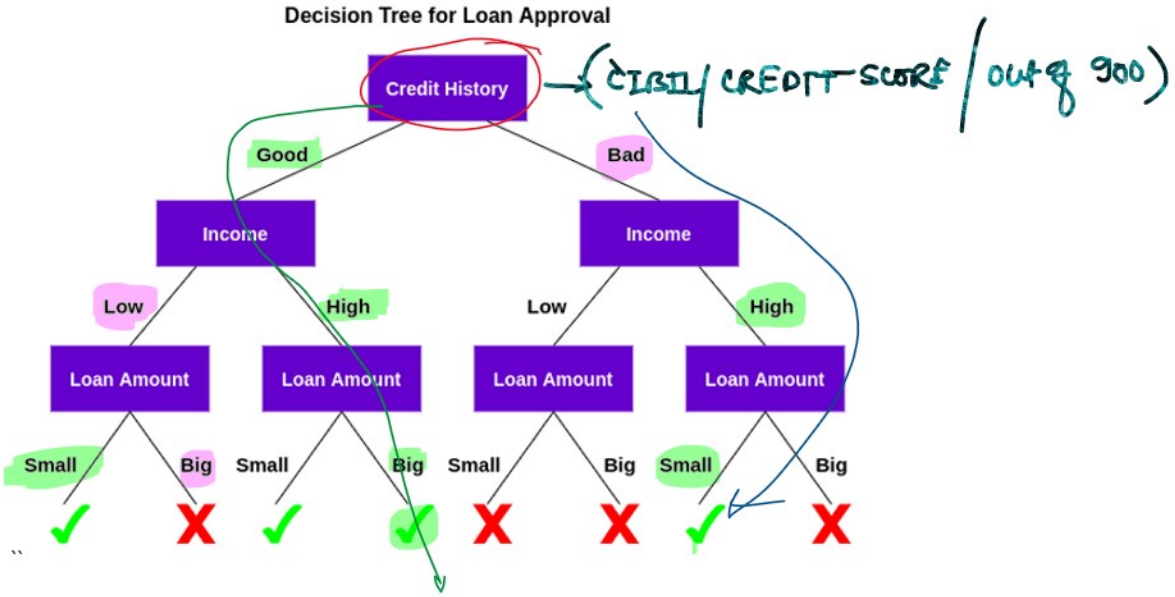
PRICE
BRAND
RAM
ROM
CPU
GPU
DISPLAY



"Buy the laptop"

"Buy the laptop"

Loan approval decision Trees



Decision trees upside down tree