Decision Trees: Study Notes

* **Definition:** A diagram showing choices and possible results to aid decision-making.
* **Structure:**
  * **Root Node:** Initial question, represents entire dataset.
  * **Branches:** Connect nodes, show decision flow.
  * **Internal Nodes:** Decision points based on features.
  * **Leaf Nodes:** Terminal nodes, represent final outcomes/predictions.
* **Types:**
  * **Classification Trees:** Predict categorical outcomes (e.g., spam/not spam).
  * **Regression Trees:** Predict continuous numerical values (e.g., house price).
* **How They Work:**
  1. Start at the root node (initial question).
  2. Ask yes/no questions to split data into subsets.
  3. Follow branches based on answers.
  4. Continue branching until a final outcome/decision is reached.
* **Advantages:**
  * Simple and easy to understand.
  * Versatile (classification and regression).
  * No feature scaling needed.
  * Handles non-linear relationships.
* **Disadvantages:**
  * Overfitting (poor performance on new data).
  * Instability (small input changes, big prediction changes).
  * Bias towards features with more levels.
* **Applications:**
  * Loan Approval in Banking

    * Medical Diagnosis
    * Predicting Exam Results in Education
* **Related Fields:**
    * Machine Learning
    * Data Mining
    * Statistics