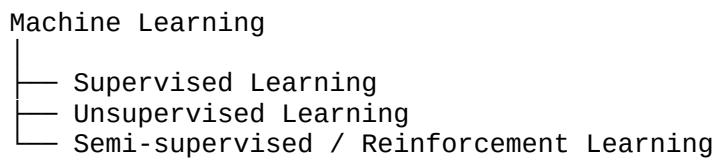


# 1 The TOP-LEVEL HIERARCHY (Start Here)

All ML models first split into **three big families** based on *learning signal*:



We'll focus mainly on the **first two**, because that's where most business ML lives.

---

## 2 SUPERVISED LEARNING (Most Business Problems)

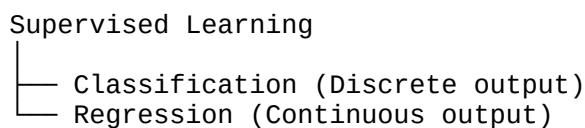
### Definition

👉 You have **input (X)** and **known output (y)**.

Examples:

- Churn (Yes/No)
  - Fraud (Yes/No)
  - Price prediction
  - Demand forecasting
- 

### 2.1 Supervised Learning → By Output Type



#### ◆ Classification (Yes / No / Categories)

Examples:

- Churn prediction
- Spam detection
- Disease detection

## ◆ Regression (Numbers)

Examples:

- Revenue prediction
  - Price prediction
  - Time-to-failure
- 

# 3 SUPERVISED MODELS → BY HOW THEY LEARN

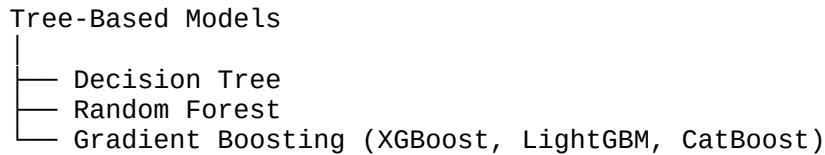
This is the hierarchy you're really asking for 

---

## 3.1 Rule-Based / Tree-Based Models (Interpretable)

### Core idea

Learn **if-else rules** from data.



### Characteristics

- Pre-hoc interpretability
- Handles non-linearities
- No scaling required
- Works very well on tabular data

### Classification / Regression

Model	Classification	Regression
Decision Tree	✓	✓
Random Forest	✓	✓
Gradient Boosting	✓	✓

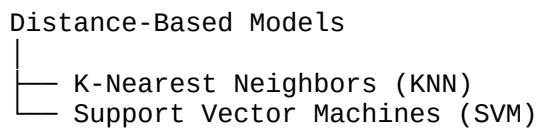
📌 Your churn project lives here.

---

## 3.2 Distance-Based Models (Geometry-Driven)

### Core idea

Similar points → similar outputs.



### Characteristics

- Require feature scaling
- Sensitive to distance metric
- Decision boundaries are geometric
- Harder to explain

### Classification / Regression

**Model Classification Regression**

KNN	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SVM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (SVR)

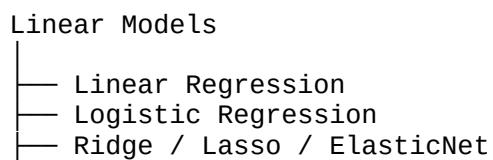
👉 Your SVM churn project lives here.

---

## 3.3 Linear / Parametric Models (Equation-Driven)

### Core idea

Learn a weighted sum of features.



### Characteristics

- Simple
- Fast
- Strong assumptions
- Easy to explain

## Classification / Regression

Model	Classification	Regression
Logistic Regression	✓	✗
Linear Regression	✗	✓
Ridge / Lasso	✗	✓

📌 Often used as **baseline models**.

---

## 3.4 Probabilistic / Bayesian Models

### Core idea

Model uncertainty explicitly.

#### Probabilistic Models

- Naive Bayes
- Bayesian Regression
- Hidden Markov Models

### Characteristics

- Fast
- Strong assumptions
- Good with small data
- Outputs probabilities naturally

## Classification / Regression

Model	Classification	Regression
Naive Bayes	✓	✗
Bayesian Regression	✗	✓

---

## 3.5 Neural Networks (Black-Box Models)

### Core idea

Learn complex non-linear functions via layers.

#### Neural Networks

- Feedforward NN
- CNN
- RNN / LSTM
- Transformers

## Characteristics

- Very powerful
- Data-hungry
- Hard to interpret
- SHAP often needed

## Classification / Regression

Model	Classification	Regression
Neural Networks	✓	✓

⚠️ Usually **not the first choice** for tabular business data.

---

# 4 UNSUPERVISED LEARNING (No Labels)

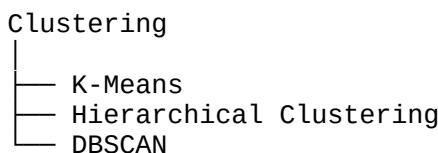
## Definition

👉 You only have X, no y.

Used for:

- Exploration
- Segmentation
- Anomaly detection
- Feature learning

## 4.1 Clustering Models



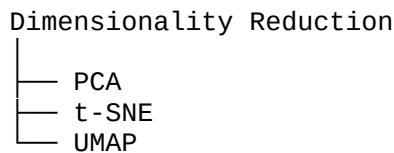
Use cases:

- Customer segmentation
- Market segmentation

⚠️ **No classification or regression here** — no labels.

---

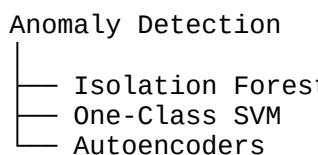
## 4.2 Dimensionality Reduction



Use cases:

- Visualization
  - Noise reduction
  - Feature compression
- 

## 4.3 Anomaly Detection



Used for:

- Fraud
  - Network intrusion
  - Rare events
- 

# 5 HYBRID / ADVANCED CATEGORIES

## Semi-Supervised Learning

- Few labels + many unlabeled points

## Reinforcement Learning

- Sequential decisions
  - Rewards
  - Not typical for churn
-

## 6 Where Your Current Knowledge Fits (Very Important)

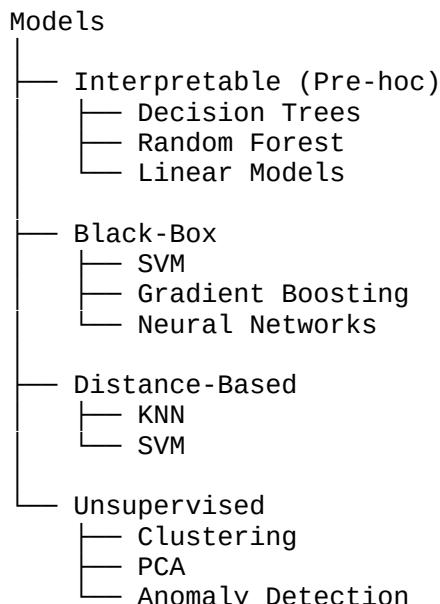
You now understand:

Model Family	You Understand
Tree-based	<input checked="" type="checkbox"/> Deep
Distance-based	<input checked="" type="checkbox"/> Deep
Cost-based decisioning	<input checked="" type="checkbox"/>
Interpretability	<input checked="" type="checkbox"/> Advanced
SHAP	<input checked="" type="checkbox"/> Advanced

That's **already strong senior-level coverage** for tabular ML.

---

## 7 Final Mental Map (One Glance)



## 8 Interview-Ready Summary (Memorize This)

"I choose models based on interpretability, data type, and business constraints. For tabular churn problems, tree-based models offer the best balance of performance and explainability, while distance-based and black-box models are considered when performance gains justify the complexity."

That's a **very strong answer**.

---

If you want next, we can:

- Build a **model-selection decision tree**
- Map **each model to real-world use cases**
- Prepare **interview Q&A on model choice**
- Compare **tree vs SVM vs NN on tabular data**

Just tell me 