□ Day 9 – Random Forest, Unsupervised Learning & Real-World Applications

---

📌 What We Covered Today

Today's session took a major leap from theoretical learning to real-world machine learning applications. Our focus was on:

✅ Customer Churn Prediction using Random Forest

✅ A beginner-friendly introduction to Unsupervised Learning

✅ Conceptual walkthrough of Heart Disease Prediction

This session showed how machine learning, especially tree-based models, is practically used in industries like telecom and healthcare.

🌲 Random Forest Recap – Strength in Numbers

Before jumping into our main project, we revisited Random Forest, an ensemble learning technique that:

Combines multiple decision trees

Uses bagging to reduce overfitting

Handles both categorical and numerical data well

Real-world tasks like churn prediction or disease diagnosis often use Random Forest for its high accuracy and reliability.

📊 Customer Churn Prediction Using Random Forest

📁 Dataset: Telco Customer Churn

This dataset records customer attributes such as:

Tenure

Contract type

Payment method

Internet service

Churn status (Yes/No – our target)

□ Data Preprocessing

To ensure clean input for the model, we performed:

Removal of rows where tenure == 0 (invalid data)

Mapped SeniorCitizen column: 0 → No, 1 → Yes

Applied Label Encoding to convert categorical columns into numerical form

```
# Map SeniorCitizen values

df["SeniorCitizen"] = df["SeniorCitizen"].map({0: "No", 1: "Yes"})

# Label encode categorical features

df = df.apply(lambda x: LabelEncoder().fit_transform(x) if x.dtype == 'object' else x)
```

▯ Model Building: Random Forest Classifier

We split the dataset (70-30) using stratified sampling to maintain class balance. Then, trained a Random Forest model with:

500 estimators (trees)

Out-of-bag evaluation

Controlled depth and features for better generalization

```
model_rf = RandomForestClassifier(

    n_estimators=500,

    oob_score=True,

    n_jobs=-1,

    random_state=50,

    max_features="sqrt",

    max_leaf_nodes=30

)
```

✅ **Results**

- 📈 **Accuracy**: ~80% on the test set
- 📊 **Confusion Matrix**: Balanced predictions across both churn and non-churn classes

The model showed solid predictive performance, reflecting real-life telecom use cases where retaining customers is critical.

▯ Unsupervised Learning – Concept Introduction

Unlike supervised learning (where target labels are known), **unsupervised learning** deals with **unlabeled data**. The goal is to discover hidden patterns or structures in the dataset.

✨ **Common Techniques:**

1. **Clustering** (e.g., K-Means, Hierarchical)
   - Used for customer segmentation, grouping behavior
2. **Dimensionality Reduction** (e.g., PCA)
   - Used for visualization, noise removal, and speeding up models

Example Applications:

- Grouping users by purchase history

- News classification without tags

- Market basket analysis

We'll explore clustering algorithms like **K-Means** in upcoming sessions.

❤ Concept: Heart Disease Prediction Using ML

We briefly explored the concept of applying ML to healthcare diagnostics, especially predicting heart disease risk.

□ How it works:

Medical features: age, cholesterol, BP, ECG results, etc.

ML models: Logistic Regression, SVM, Random Forest

Output: Predicts risk status (0 = No risk, 1 = At risk)

Even though we didn't code it today, the discussion emphasized how machine learning is revolutionizing healthcare by assisting doctors with quick, data-driven decisions.

📌 Key Takeaways
✅ Trained and evaluated a real-world Customer Churn model using Random Forest
✅ Practiced data cleaning, label encoding, and performance evaluation
✅ Learned the theory and examples of Unsupervised Learning
✅ Understood how ML supports critical industries like telecom and healthcare

---

📝 **Conclusion**

Today's session helped bridge the gap between theoretical ML and its **real-world implementation**. We not only built a working churn prediction model but also gained valuable insight into **unsupervised learning** and **healthcare applications**. With each session, we're now moving closer to using ML in practical, impactful ways.