

AI & ML Internship Task 5: Decision Trees and Random Forests

Objective:

Learn tree-based models for classification using the Titanic dataset.

Dataset:

Titanic survival dataset with features like Pclass, Sex, Age, SibSp, Parch, Fare, and Embarked.

Preprocessing:

- Removed irrelevant columns: PassengerId, Name, Ticket, Cabin
- Filled missing values in Age (median) and Embarked (mode)
- Encoded categorical variables (Sex, Embarked)

Model Training:

1. Decision Tree Classifier:

- Accuracy: 0.78
- Visualized tree using Scikit-learn's `export_text`

2. Random Forest Classifier:

- Accuracy: 0.82
- Used to derive feature importance

Top Feature Importances (Random Forest):

Sex 0.271410

Fare 0.265010

Age 0.249995

Pclass 0.086957
SibSp 0.053685
Parch 0.039897
Embarked 0.033044

Partial Decision Tree Rules:

|--- Sex <= 0.50
| |--- Pclass <= 2.50
| | |--- Age <= 2.50
| | | |--- Parch <= 1.50
| | | | |--- class: 1
| | | |--- Parch > 1.50
| | | | |--- class: 0
| | |--- Age > 2.50
| | | |--- Age <= 27.50
| | | | |--- Age <= 24.50
| | | | | |--- Fare <= 13.40
| | | | | | |--- Age <= 22.50
| | | | | | | |--- class: 1
| | | | | | |--- Age > 22.50
| | | | | | | |--- class: 0
| | | | | |--- Fare > 13.40
| | | | | |--- class: 1
| | | |--- Age > 24.50
| | | | |--- Parch <= 1.50
| | | | | |--- Fare <= 17.43
| | | | | | |--- class: 1

| | | | | | | |--- Fare > 17.43

| | | | | | | |--- Age <= 26.00

| | | | | | | |--- class: 1

| | | | | | | |--- Age > 26.00

| | | | | | | |--- class: 0

| | | | | |--- Parch > 1.50

| | | | ...