# Forest Cover Type Prediction Report

## Objective

Build a system that can predict the type of forest cover using analysis data for a 30m x 30m patch of land in the forest.

## Dataset

This dataset is an analysis dataset from the forest department performed in the Roosevelt National Forest of northern Colorado.

### Forest Cover Types (Integer Classification)

| Label | Forest Cover Type |
|---|---|
| 1 | Spruce/Fir |
| 2 | Lodgepole Pine |
| 3 | Ponderosa Pine |
| 4 | Cottonwood/Willow |
| 5 | Aspen |
| 6 | Douglas-fir |
| 7 | Krummholz |

### Main Features

- **Elevation**: Elevation in meters
- **Aspect**: Aspect in degrees azimuth
- **Slope**: Slope in degrees
- **Horizontal_Distance_To_Hydrology**: Horizontal distance to nearest surface water features
- **Vertical_Distance_To_Hydrology**: Vertical distance to nearest surface water features
- **Horizontal_Distance_To_Roadways**: Horizontal distance to nearest roadway
- **Hillshade_9am**: Hillshade index at 9am (0 to 255)
- **Hillshade_Noon**: Hillshade index at noon (0 to 255)
- **Hillshade_3pm**: Hillshade index at 3pm (0 to 255)
- **Horizontal_Distance_To_Fire_Points**: Horizontal distance to nearest wildfire ignition points
- **Wilderness_Area**: 4 binary columns (0 = absence, 1 = presence)
- **Soil_Type**: 40 binary columns (0 = absence, 1 = presence)
- **Cover_Type**: Target variable (forest cover type)

## Model

A Random Forest Classifier was used, with hyperparameter tuning via GridSearchCV. The best model was selected based on cross-validation accuracy.

### Best Parameters

- `max_depth`: None
- `min_samples_split`: 2
- `n_estimators`: 200

## Results

### Test Accuracy

- **0.87**

### Classification Report

| Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 1 | 0.78 | 0.78 | 0.78 | 432 |
| 2 | 0.80 | 0.66 | 0.72 | 432 |
| 3 | 0.86 | 0.82 | 0.84 | 432 |
| 4 | 0.94 | 0.98 | 0.96 | 432 |
| 5 | 0.89 | 0.95 | 0.92 | 432 |
| 6 | 0.84 | 0.90 | 0.87 | 432 |
| 7 | 0.94 | 0.97 | 0.95 | 432 |

| Metric | Value |
|---|---|
| Accuracy | 0.87 |
| Macro Avg F1 | 0.86 |
| Weighted Avg F1 | 0.86 |

## Model Persistence

The best model was saved to disk as `forest_cover_model.joblib` for future use.