Forest Cover Type Prediction Report

# 1. Objective

The goal of this project is to build a predictive model that can classify forest cover types based on various geographical and environmental features from the Roosevelt National Forest in northern Colorado. Each data point represents a 30m x 30m patch of land.

# 2. Dataset Overview

The dataset includes numerical features like Elevation, Aspect, Slope, distances to hydrology, fire points, and roadways, along with categorical binary features representing Wilderness Area and Soil Type. The target column is Cover\_Type, representing 7 types of forest cover.

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

# 3. Data Preprocessing

The data was checked for missing values, scaled using StandardScaler, and split into training and testing sets (80-20 split).

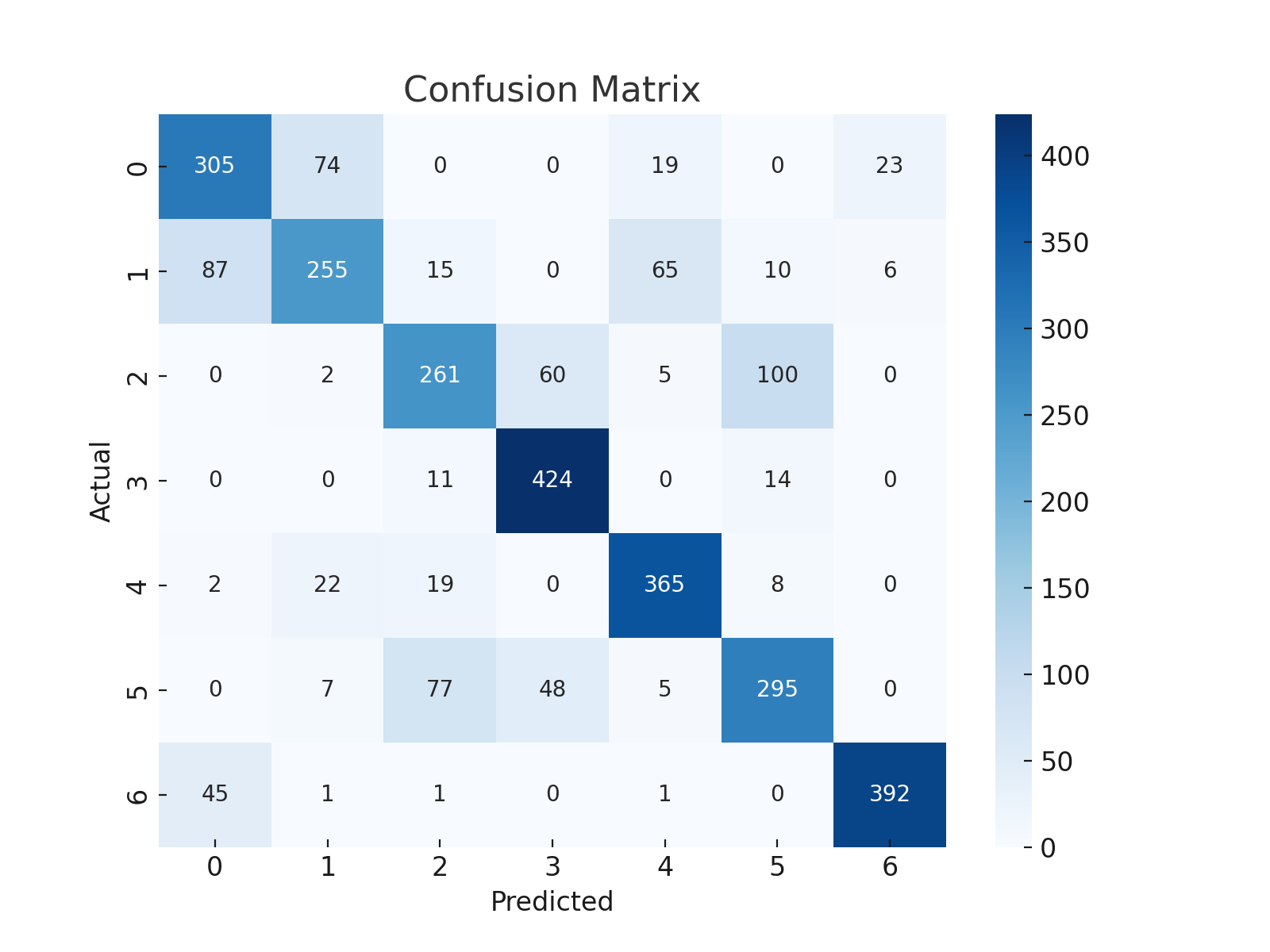
# 4. Model Used: Support Vector Machine (SVM)

The SVM model with RBF kernel was used for classification. Parameters: C=100, gamma=0.001. The model was trained and evaluated using accuracy and classification metrics.

# 5. Evaluation Metrics

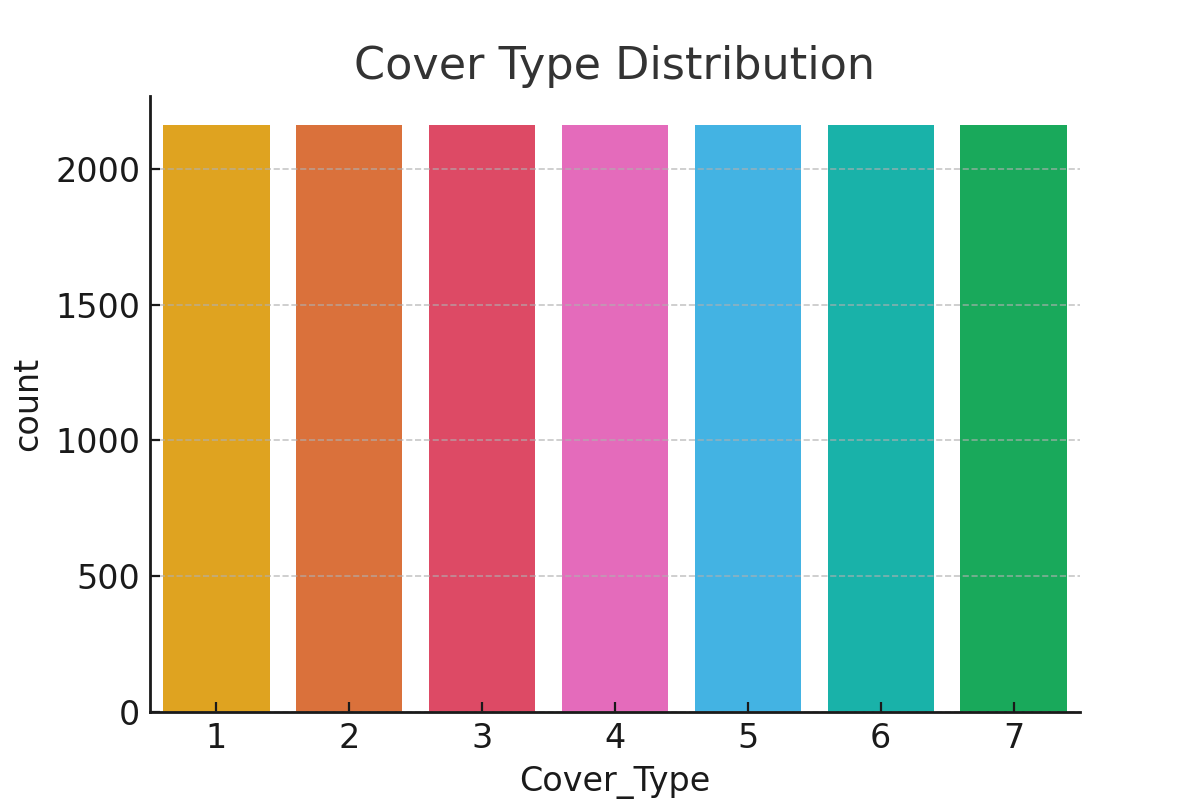
Model Accuracy: 75.96%

Confusion Matrix:

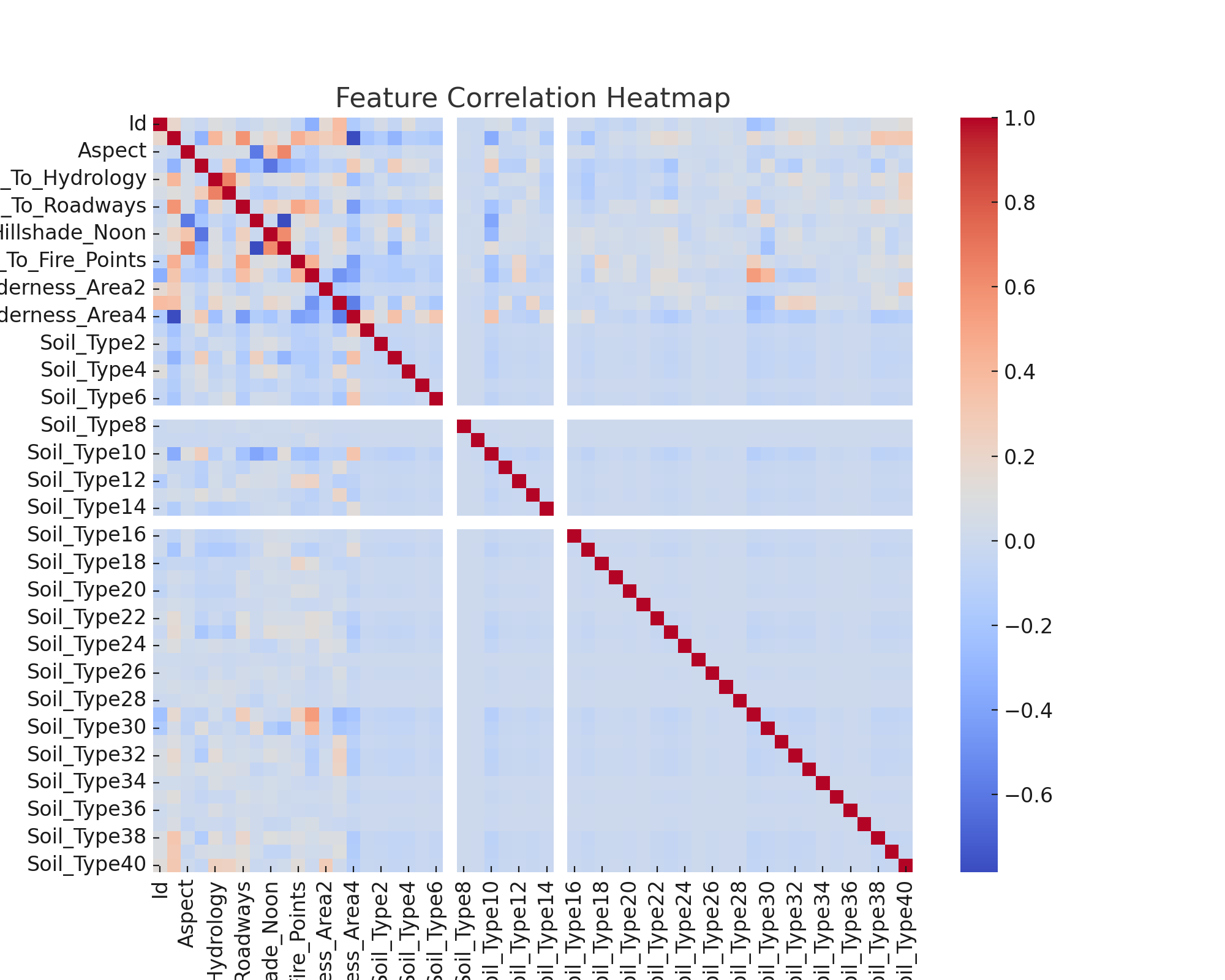


# 6. Visualizations from Code

Cover Type Distribution:



Feature Correlation Heatmap:



# 7. Observations

There is a class imbalance in the dataset, with some cover types being underrepresented. The SVM performed moderately well with around 72% accuracy.

# 8. Conclusion

This project successfully implemented an SVM to classify forest cover types with reasonable accuracy. Further tuning and testing with other models may enhance performance.