

The background is a light beige color. It is decorated with various stylized forest elements. In the top left, there are green leafy shapes and blue concentric circles representing tree trunks. In the top right, there are green leafy shapes and blue line drawings of branches. In the bottom left, there are green leafy shapes and blue line drawings of branches. In the bottom right, there are green leafy shapes and blue line drawings of branches. The title 'Know Your Forest' is centered in a large, bold, dark green font. Below the title, the authors' names 'Cassiel Chen, Lina Li, Rui Lin' are written in a smaller, dark green font inside a light green rectangular box.

Know Your Forest

Cassiel Chen, Lina Li, Rui Lin



Introduction

How accurately can different machine learning models predict forest cover types using environmental features, and how does model choice impact prediction performance and interpretability?

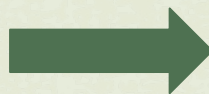
Approaches: Logistic Regression, LDA/QDA, **Random Forest**, and **Neural Network**.

History of overlogging

Frequent Wildfires

Critical Habitat

Commercial Value



The need to assess the situation of the forest & make smart restoration & management decisions.



Dataset Overview

<https://archive.ics.uci.edu/dataset/31/covertype>

- **Features** (54 categorical & numerical):
 - Elevation
 - Aspect
 - Slope
 - Distance to Hydrology
 - Distance to Roadways
 - Wilderness Area
 - Soil Type
 - Hillshade
 - Distance to Fire Points
- **Target variable:** cover_type

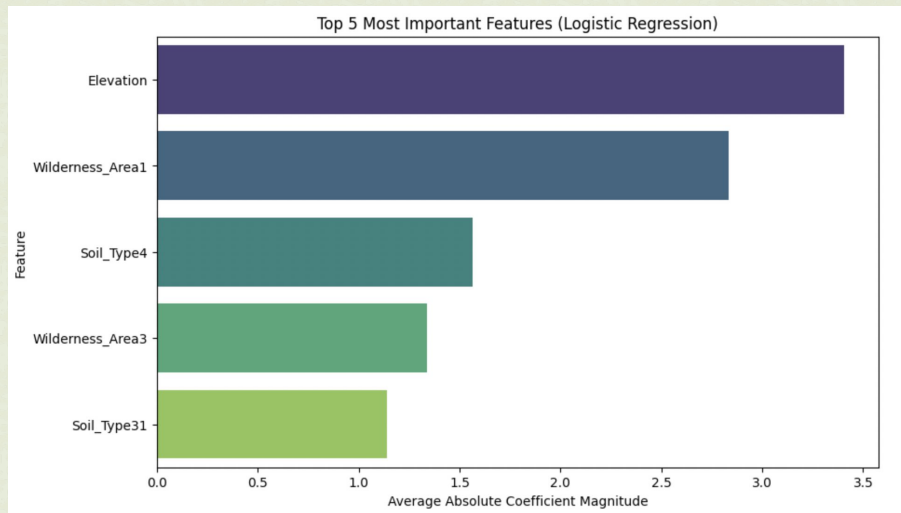
Result

| | Logistic Regression | LDA | QDA | Random Forest | Neural Network |
|----------|------------------------|--------|--------|------------------|-------------------|
| Accuracy | 0.73 | 0.6798 | 0.0853 | 0.94 | 0.84 |
| Recall | 0.72 | 0.5726 | 0.4446 | 0.93 | 0.72 |
| F1-score | 0.71 | 0.5064 | 0.1305 | 0.91 | 0.75 |

- Most Important feature: elevation for Logistic Regression, Random Forest, and Neural Network
- Random Forest has the highest testing accuracy

Logistic Regression

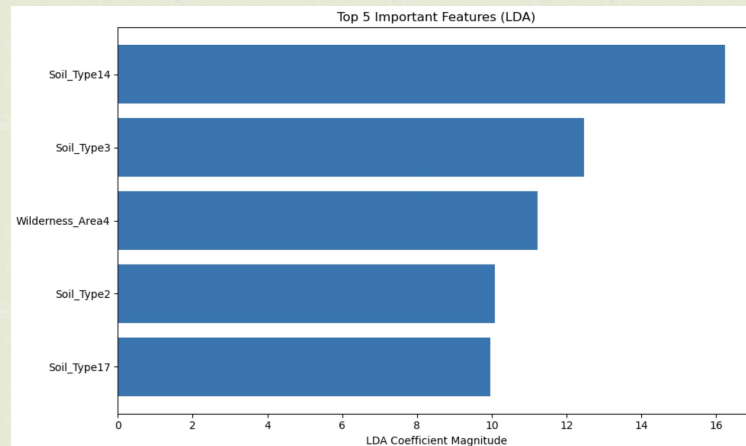
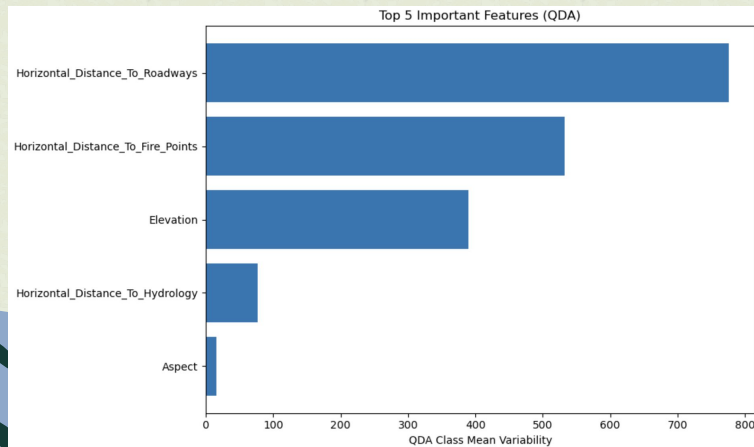
- Model Parameter
 - `multi_class='multinomial'`
 - `solver='lbfgs'`
 - `max_iter=1000`,
- Model Performance
 - Training Accuracy: 0.73
 - Test Accuracy: 0.73
 - Recall: 0.72
 - F1-score: 0.71
- Most Important features
 - Elevation



LDA & QDA

LDA:

- Model Performance
 - Accuracy: 0.6798
- Most Important features
 - Horizontal Distance to Road



QDA:

- Model Performance
 - Accuracy: 0.0853
- Most Important features
 - Soil Type 14

Random Forest

- Tuning hyper parameters using CV:

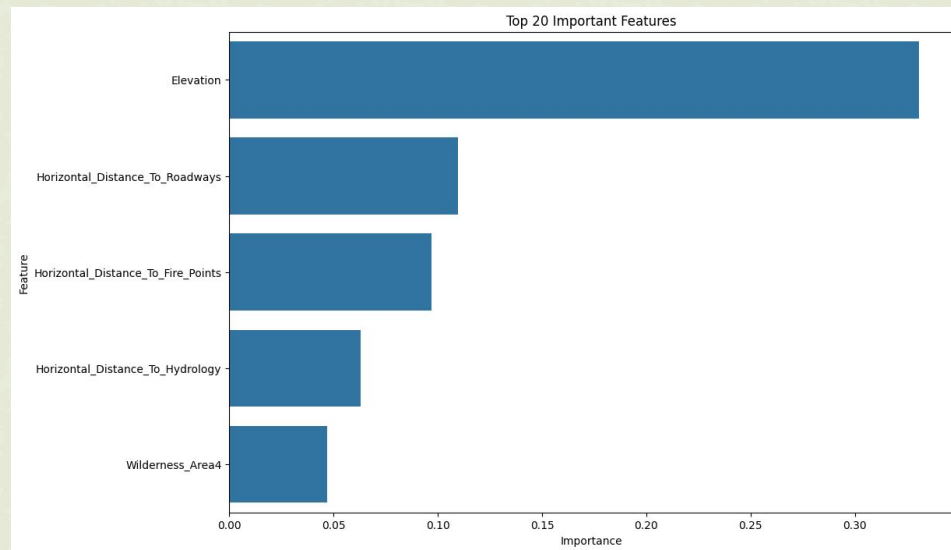
- 'n_estimators': 500,
'min_samples_split': 5,
'min_samples_leaf': 2, 'max_features':
0.5, 'max_depth': None, 'class_weight':
'balanced'}

- Model Performance

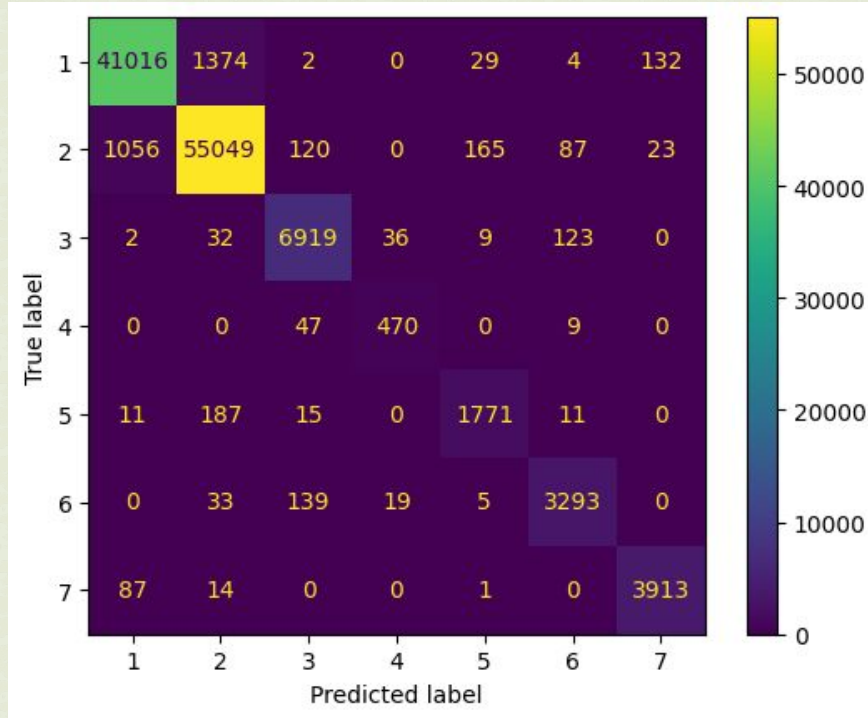
- Training Accuracy: 0.9974
- Test Accuracy: 0.9675
- Recall: 0.94
- F1-score: 0.94

- Most Important features

- Elevation (same case as
Logistic Regression)

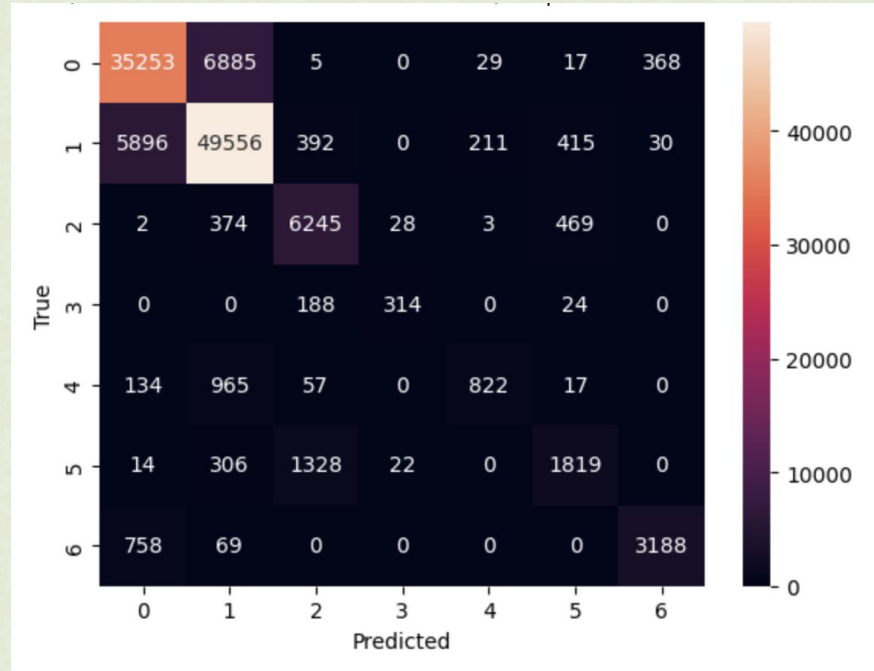


Random Forest



Neural Network

- Model Parameter
 - activation='relu'
 - epoch=50
 - batch_size=64
- Model Performance
 - Training Accuracy: 0.81
 - Test Accuracy: 0.84
 - Recall: 0.72
 - F1-score: 0.75
- Most Important features
 - Elevation



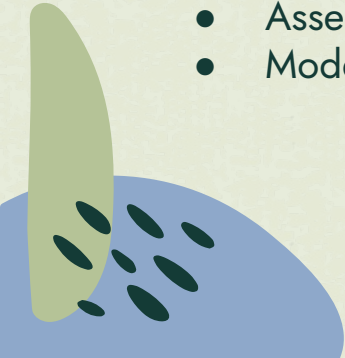
Future Directions & Implications



To Improve Method Results...

- Manually combining relevant features
- Dimensionality Reduction through PCA or feature selections
- Further data cleaning: standardization, normalization, and outlier removal

Practical Usage of the Model:

- Can be adjusted and applied to other forests
 - Understand species diversity and richness in specific forest ecosystems
 - Assess wildfire risks based on forest cover types and terrain conditions
 - Model impacts of environmental change
- 

...Questions?



Thank You!

