

Predicting Forest Cover Type

Team Checkpoint 1

Dong Bing, Richard Gunawardene, Kent Merdes, Christina Macholan, Tyler Wintermeyer

1. Introduction

Four wilderness areas: Rawah, Neota, Comanche Peak, Cache la Poudre.

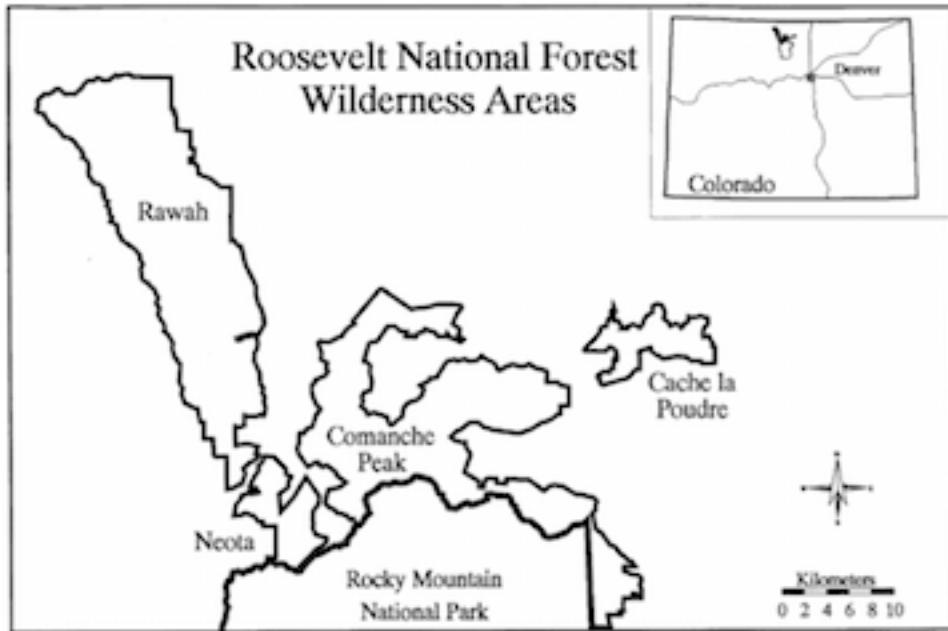


Figure 1: Map of wilderness areas included in study

2. The Modeling Problem

Our aim is to correctly classify the type of forest cover for a particular region based on

3. The Data

581,012 records 55 variables

Variable Description	Type & Measurement
Elevation	numeric in meters
Aspect	numeric in degrees azimuth
Slope	numeric in degrees
Horizontal distance to nearest surface water features	numeric in meters
Vertical distance to nearest surface water features	numeric in meters
Horizontal distance to nearest roadway	numeric in meters

Variable Description	Type & Measurement
Hillshade index at 9am during summer solstice	numeric as index (0 to 255)
Hillshade index at Noon during summer solstice	numeric as index (0 to 255)
Hillshade index at 3pm during summer solstice	numeric as index (0 to 255)
Horizontal distance to nearest wildfire ignition points	numeric in meters
Wilderness Areas (4 areas)	binary for each area (0 or 1)
- 1 - <i>Rawah</i>	
- 2 - <i>Neota</i>	
- 3 - <i>Comanche Peak</i>	
- 4 - <i>Cache la Poudre</i>	
Soil Type (40 types)	binary for each type (0 or 1)
- See Appendix A for details	
Forest Cover Type (7 types)	integer for each type (0 or 1)
- 1 - <i>Spruce/Fir</i>	
- 2 - <i>Lodgepole Pine</i>	
- 3 - <i>Ponderosa Pine</i>	
- 4 - <i>Cottonwood/Willow</i>	
- 5 - <i>Aspen</i>	
- 6 - <i>Douglas-fir</i>	
- 7 - <i>Krummholz</i>	

4. Exploratory Data Analysis (EDA)

Traditional EDA

Model-Based EDA

5. Predictive Modeling: Methods and Results

Train / Test Data

Individual Model A

Individual Model B

6. Comparison of Results

7. Conclusions

8. Bibliography

9. Appendices

```
densityplot(~covtype$Elevation|covtype$Cover_Type)
```

