

# Predicting Forest Cover Type

---

Lucy Hayes

# Obtaining Data

- UCI Machine Learning Repository
- Roosevelt Forest - CO, USA
- Cover Type, Wilderness Type

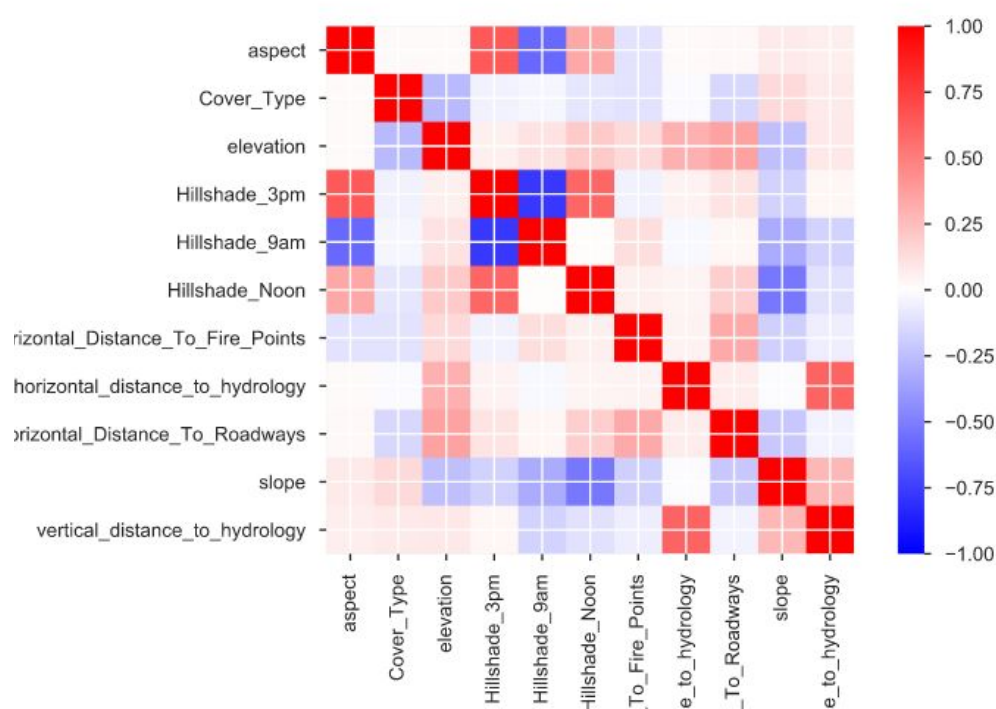
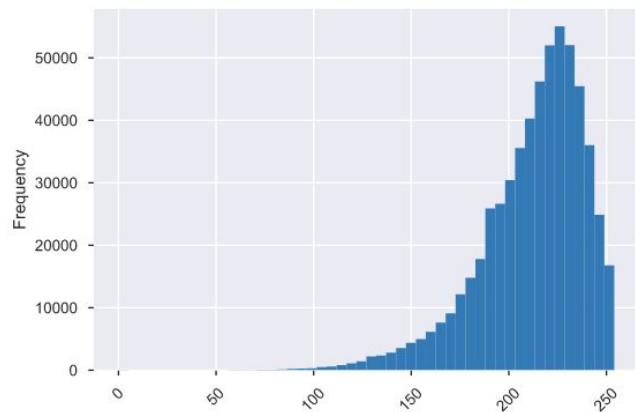


# Scrubbing Dataset

- Rename Columns
- Downsizing Dataset
- NAs and Empty Variables

# Exploring Data

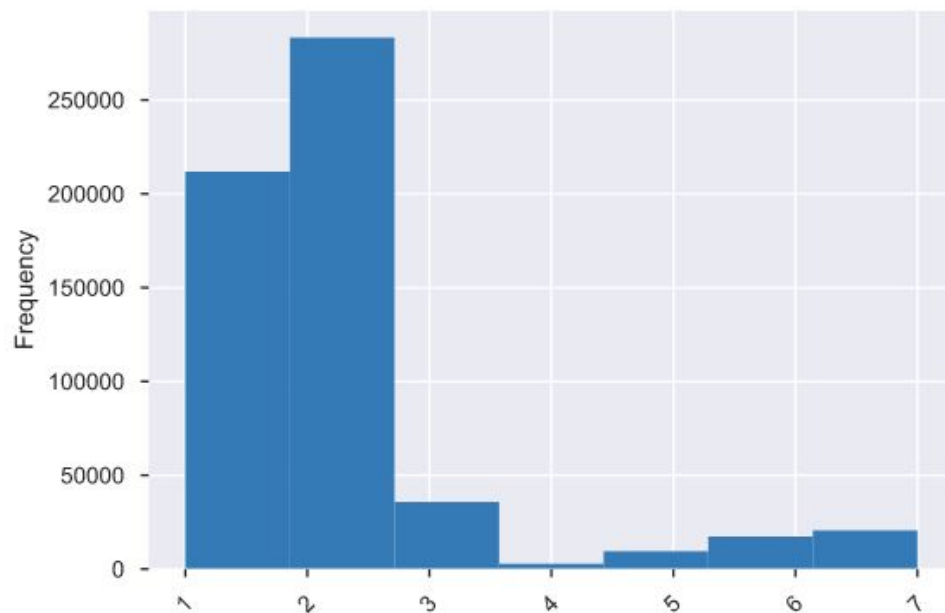
- Skew
- Collinearity



# Explore

- Low Data for Soil Types 4, 5, 6, 7.
- Sample of 50000 for computational efficiency

Cover Type	
1	18077
2	23518
3	3125
4	244
5	836
6	1466
7	1734



# Model

- Decision Tree
- Random Forest
- GridSearchCV



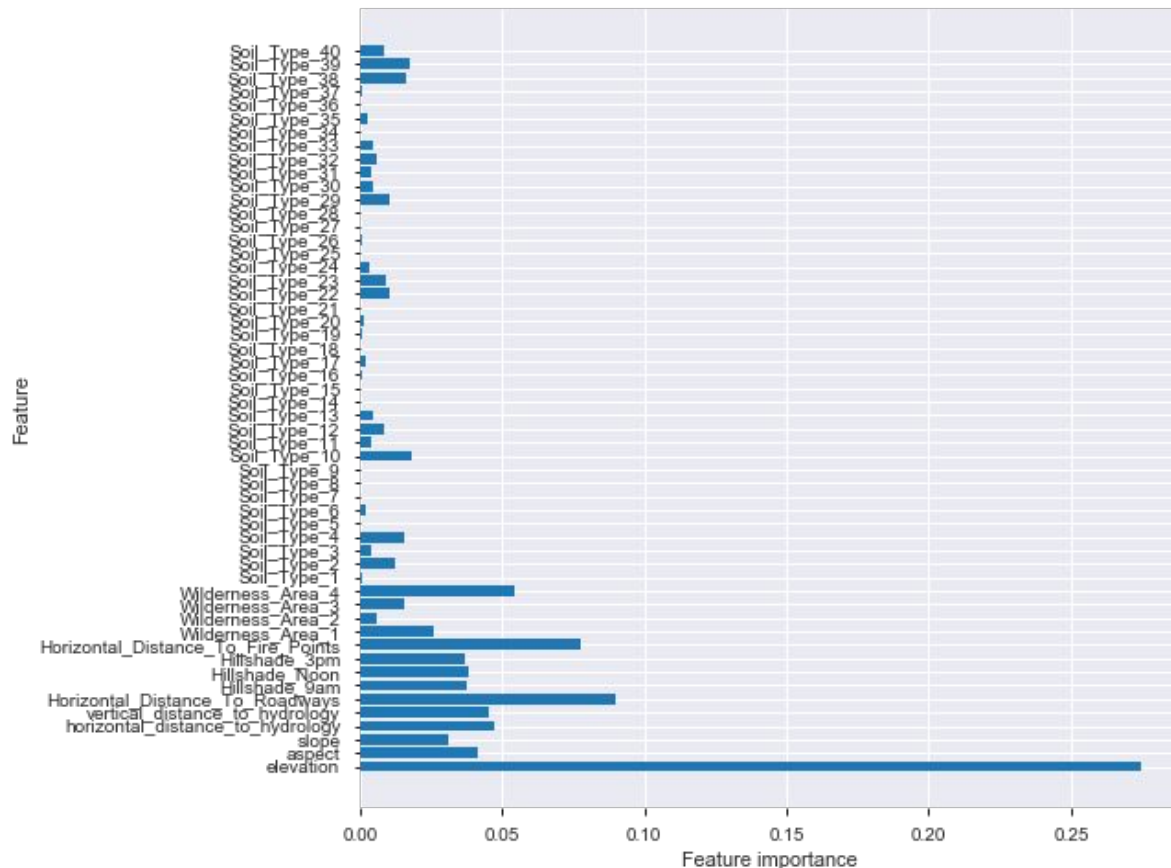
# Interpret

- Decision Tree: 82%
- Random Forest: 86%

## Features:

- Elevation vs. Soil Type

## Predictions vs Outcomes





# Future Considerations

- Using Larger Dataset
- Adjusting for Normality/Skew and Collinearity





# Questions

---