

A wide-angle photograph of a mountainous landscape. In the foreground, a steep hillside covered in green grass and small rocks slopes down towards the bottom left. In the middle ground, more mountain ridges are visible, some with patches of green vegetation and others appearing as rocky outcrops. The background shows a range of mountains under a sky filled with large, white, billowing clouds.

Data Science Project 7

ROOSEVELT NATIONAL FOREST TREE TYPE CLASSIFICATION

BY CATHY KAM
01/01/2023

The aim of this project is to:

- (1) conduct EDA based how the environmental factors correlated with the tree types.*
- (2) predict the Cover Type (565892 observations) using Random Forest Classifier, Extra Trees Classifier and Support Vector Machine.*

Then, I'll evaluate which model works the best for the classification.



ABOUT THE DATA:

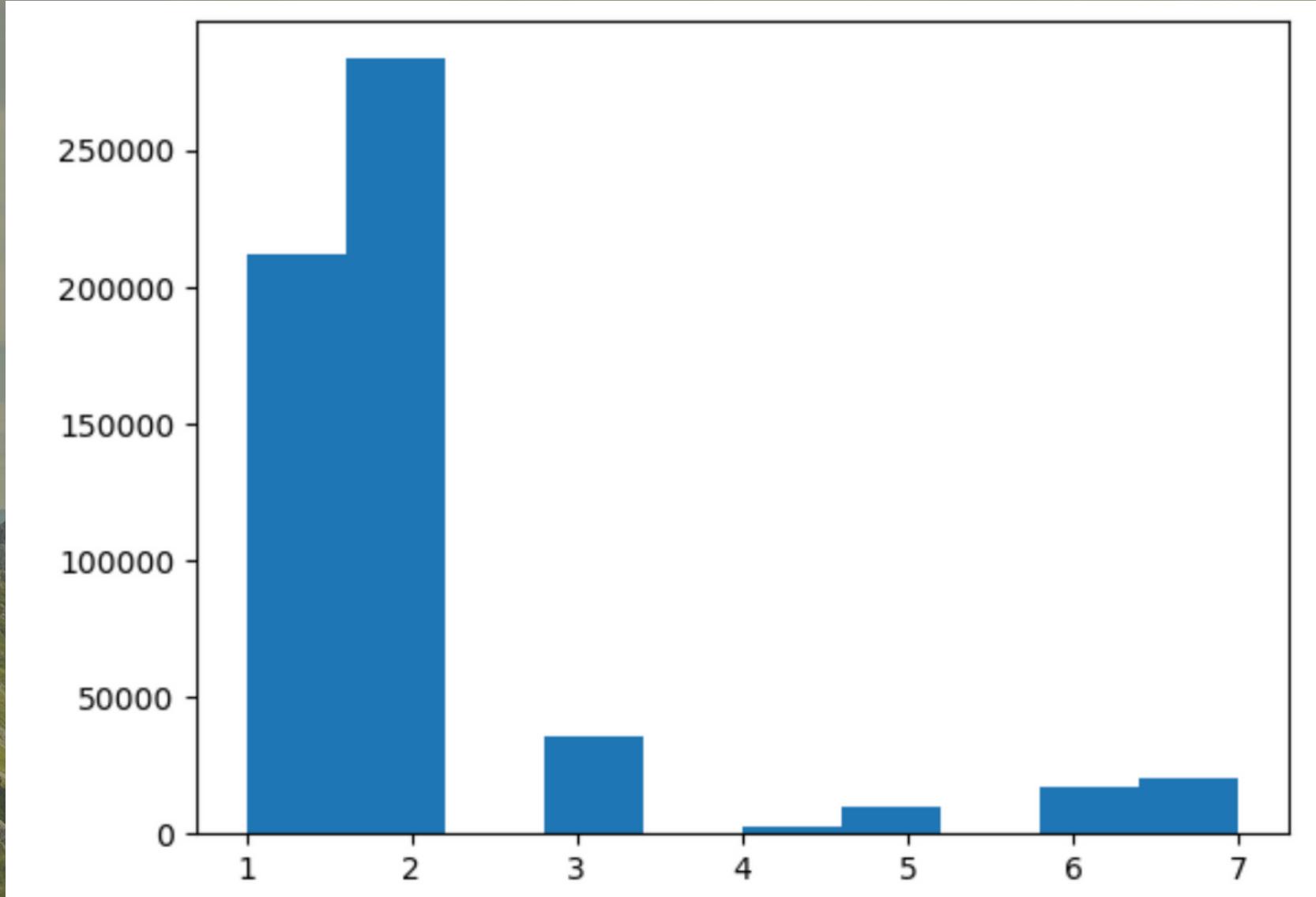
This dataset contains tree observations from four areas of the Roosevelt National Forest in Colorado. All observations are cartographic variables (no remote sensing) from 30 meter x 30 meter sections of forest.

The original database owners are Jock A. Blackard, Dr. Denis J. Dean, and Dr. Charles W. Anderson of the Remote Sensing and GIS Program at Colorado State University.

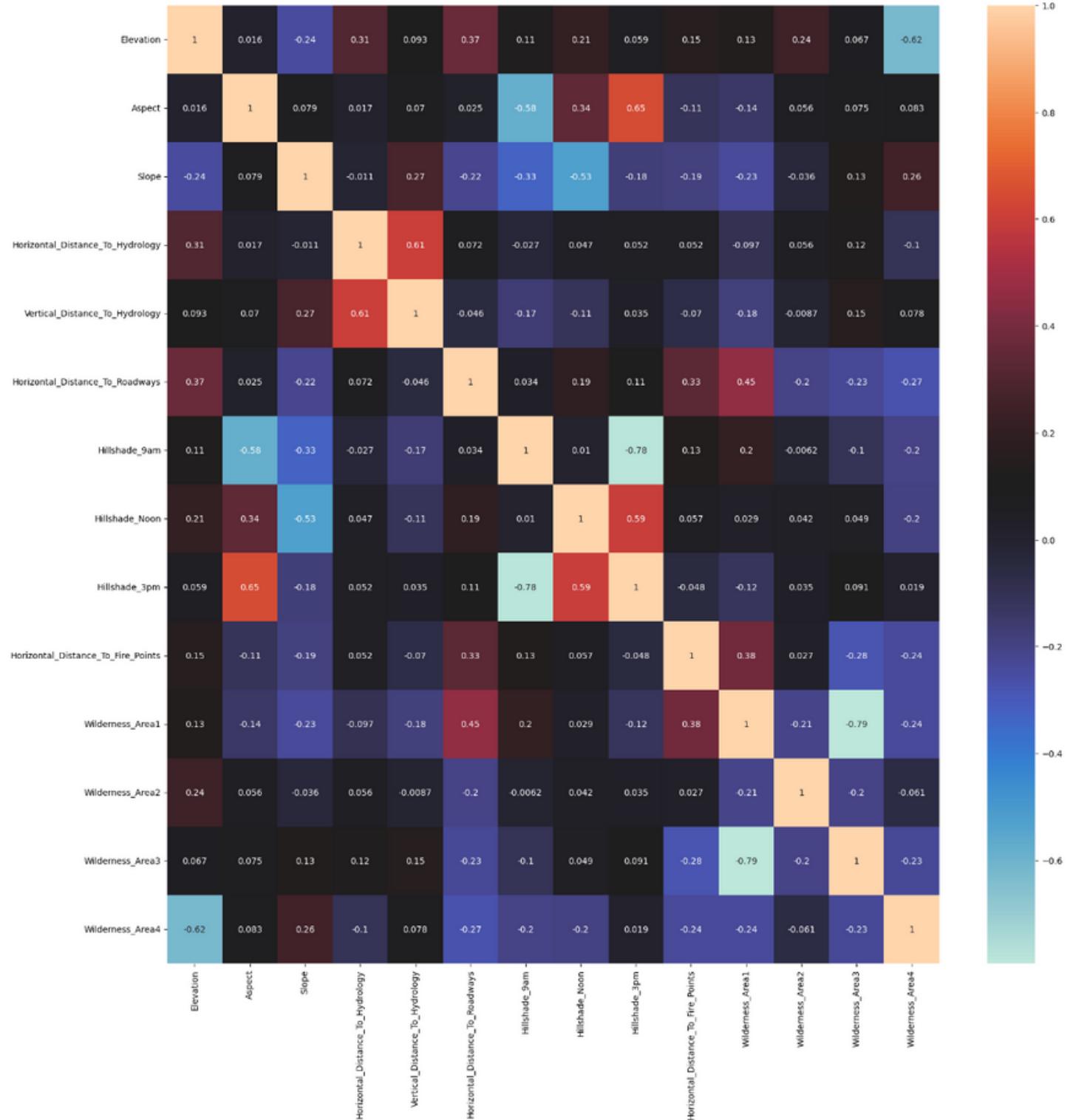


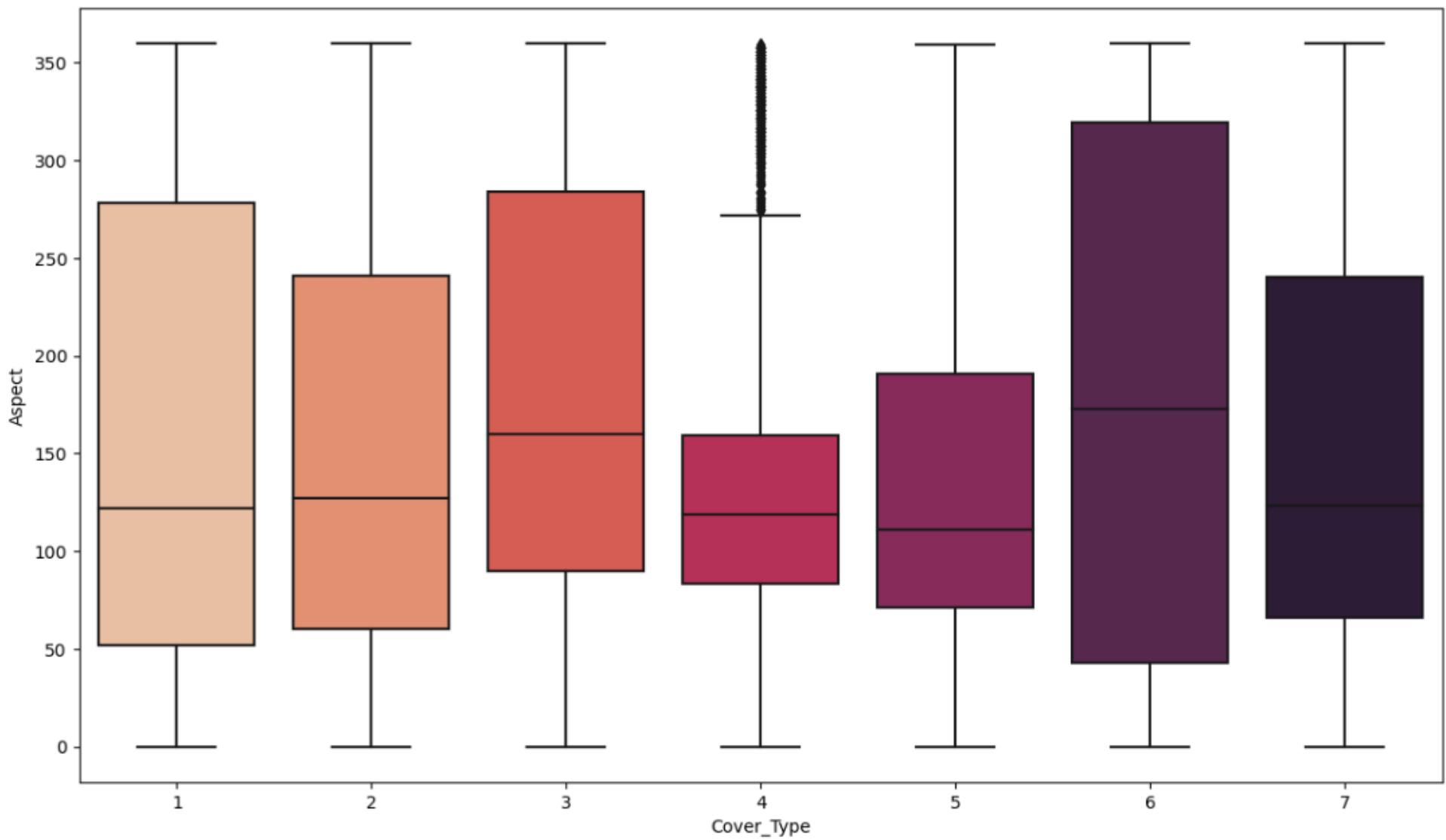
QUESTIONS TO BE ADDRESSED:

1. Is there a model that predicts what types of trees grow in an area based on the surrounding characteristics?
2. What kinds of trees are most common in the Roosevelt National Forest?
3. Which tree types can grow in more diverse environments? Are there certain tree types that are sensitive to an environmental factor, such as elevation or soil type?



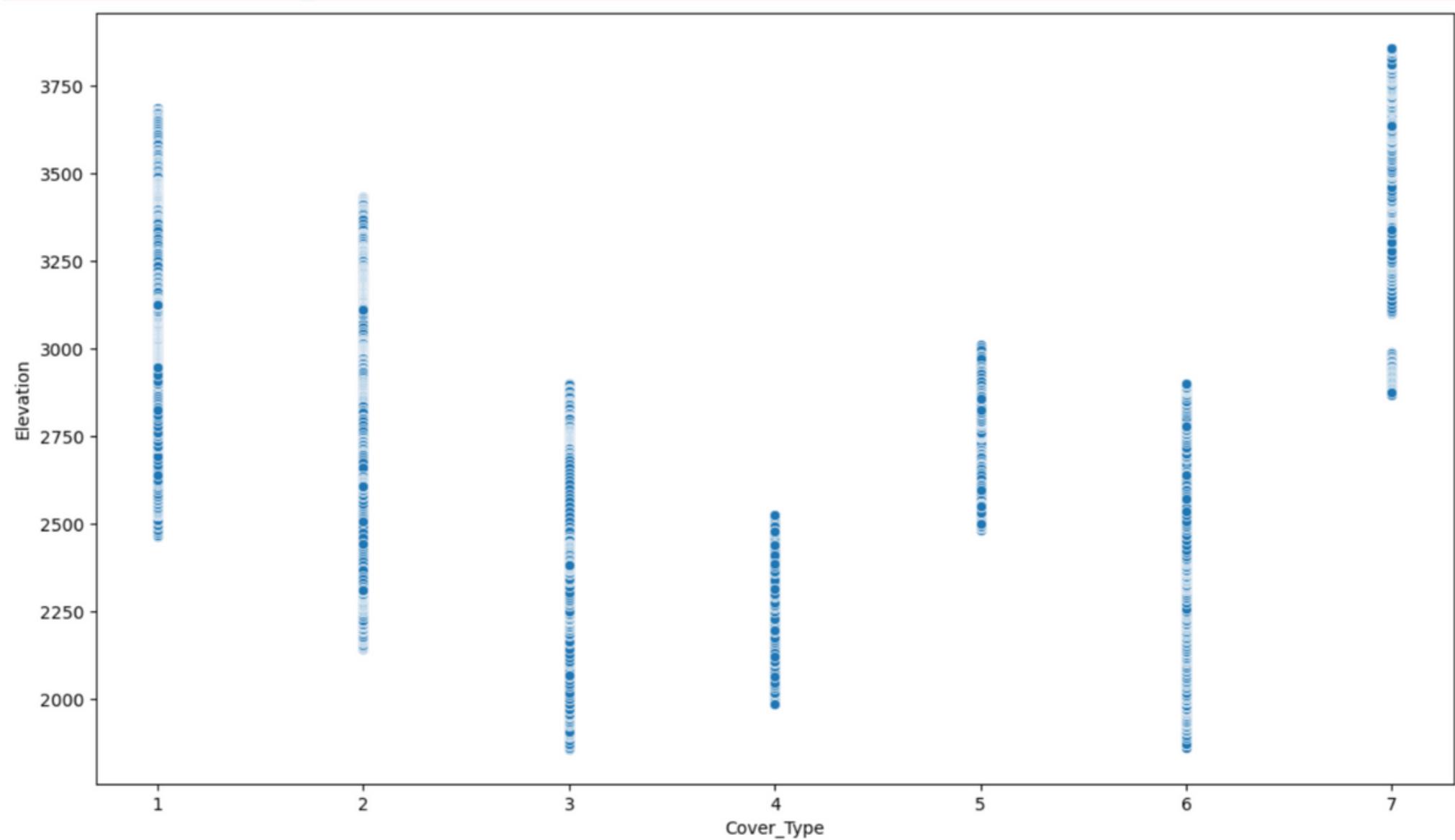
Type 2 (Lodgepole Pine) is the most common type of trees in Roosevelt National Forest



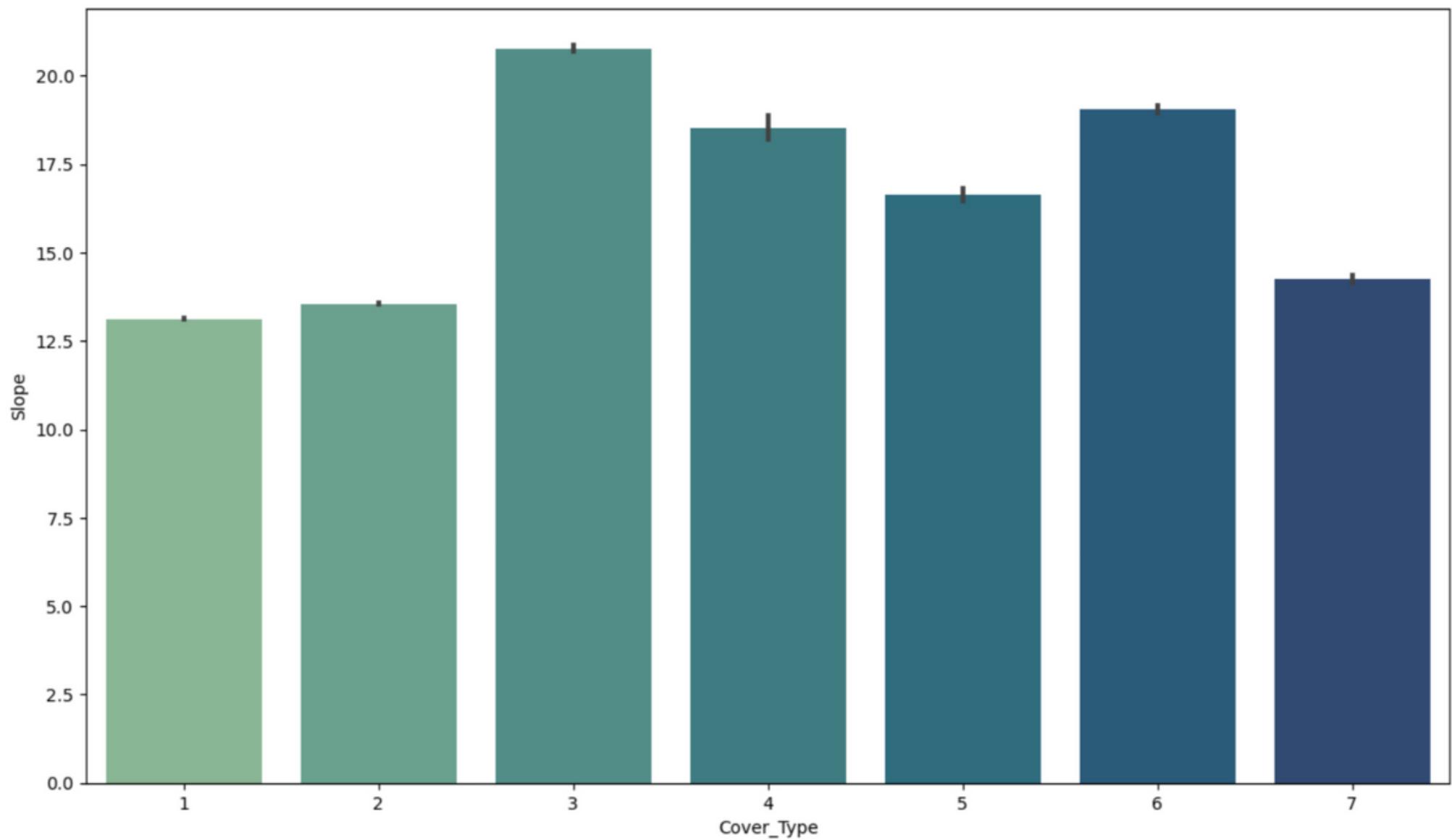


Douglas-fir (#6) has a normal distribution in terms of aspects.

While Cottonwood/Willow (#4) has a left-skewed distribution, where a significant number of outliers in aspects, ranging from 270-350.

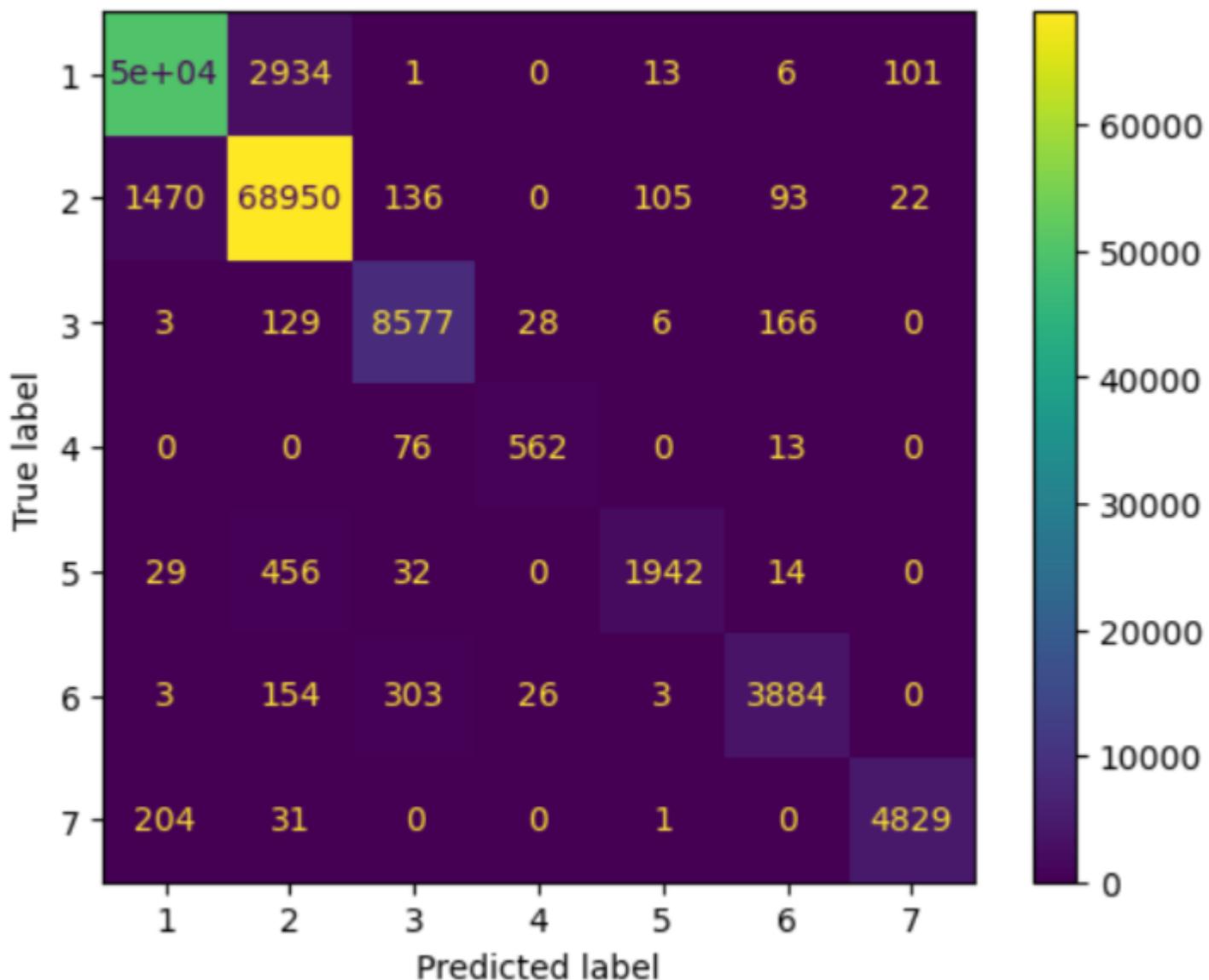


Krummholz tends to grow at a higher height, while
Cottonwood/Willow tends to grow at lower height.



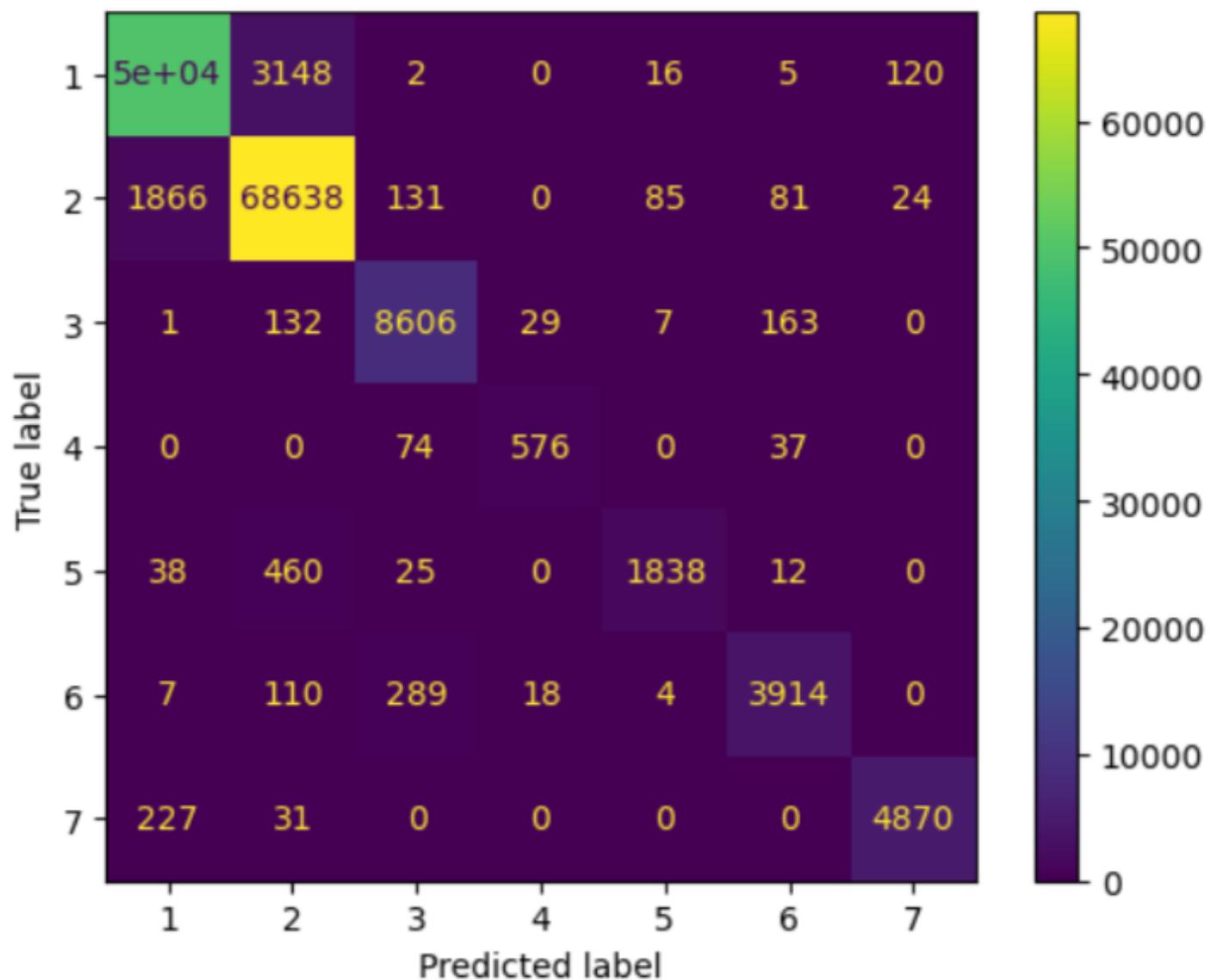
Ponderosa Pine tends to grow at steeper soil, while Spruce/Fir tends to grow at flatter soil.

RANDOM FOREST CLASSIFIER



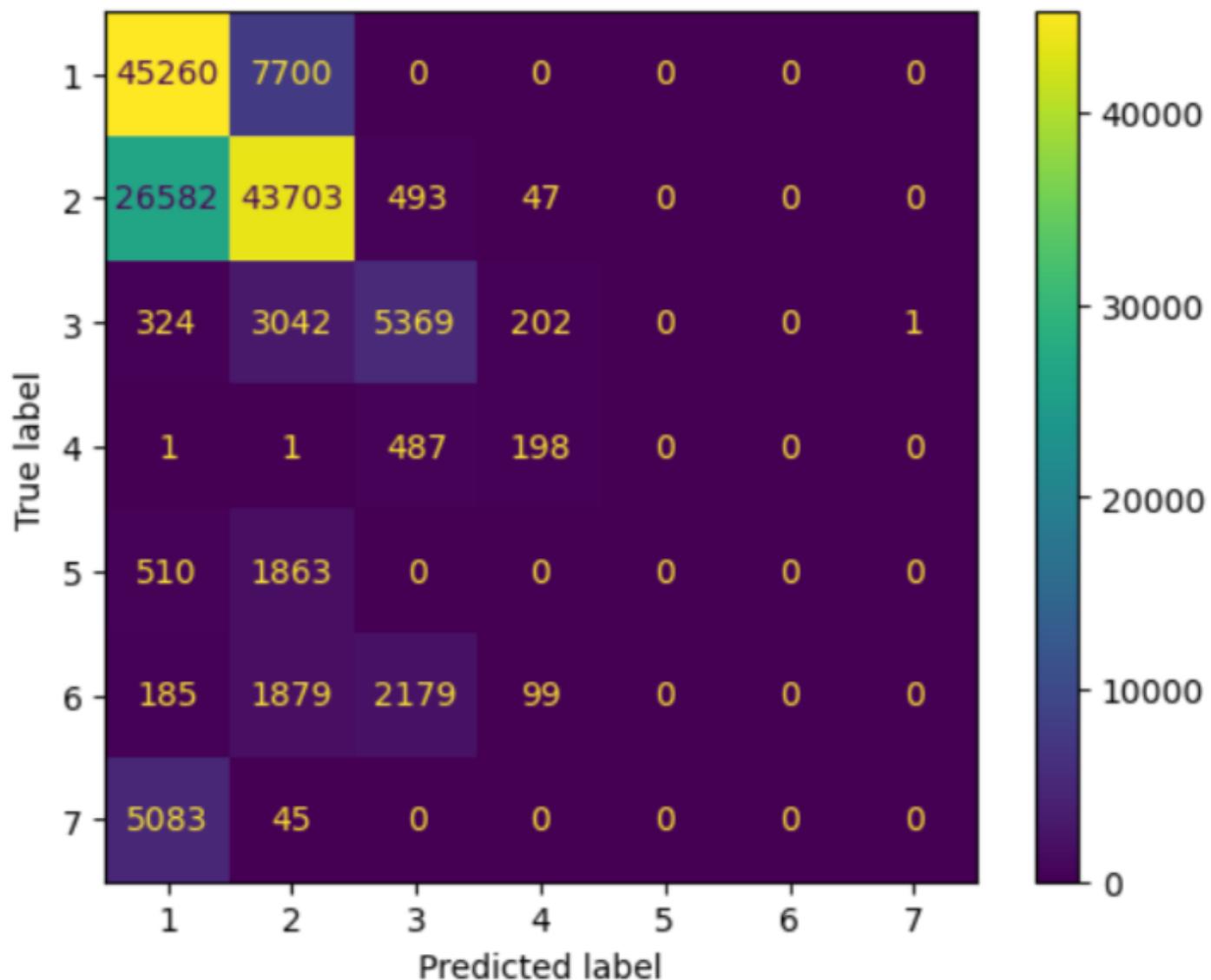
Overall test score of 95%

EXTRA TREES CLASSIFIER



Overall test score of 95%

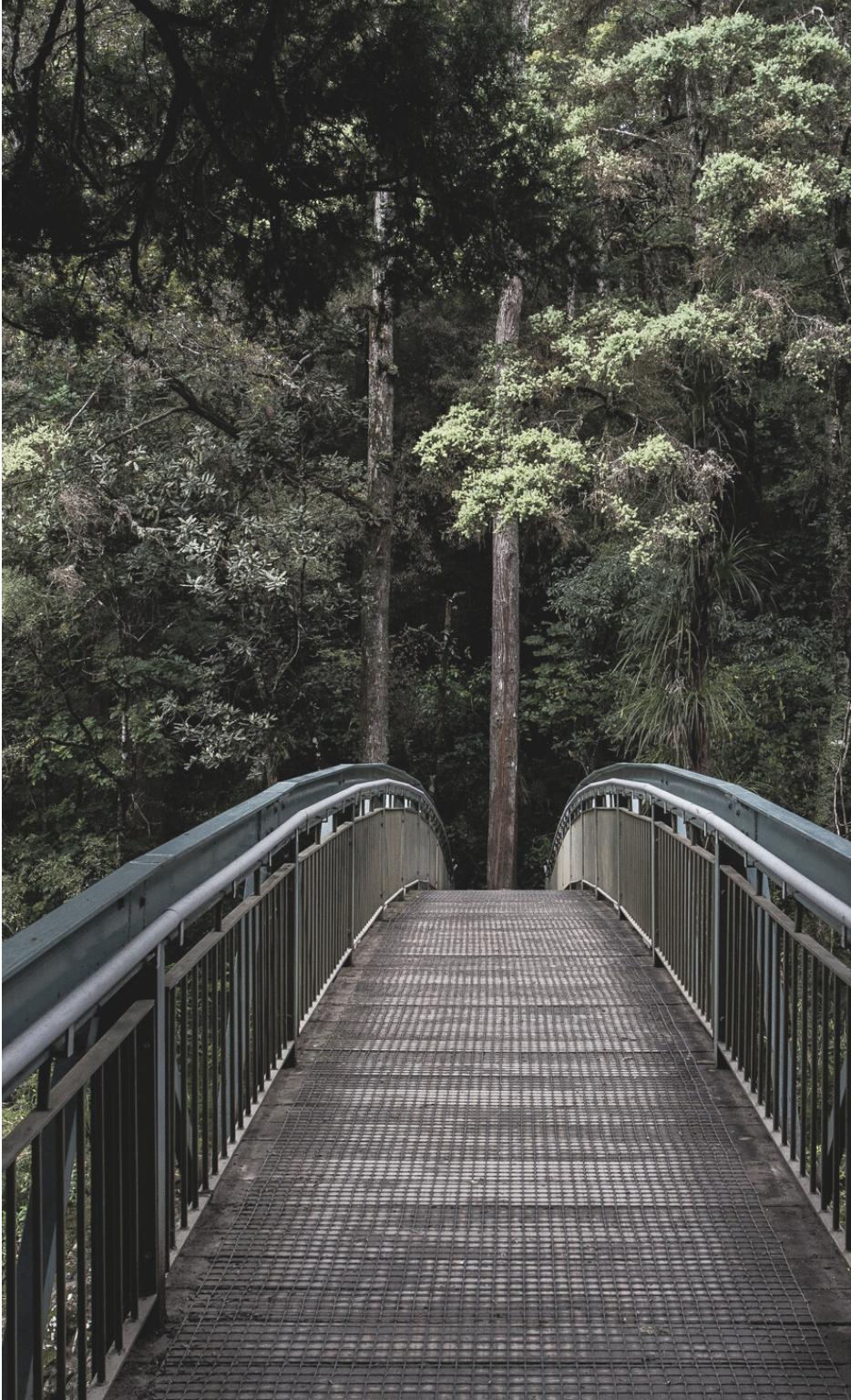
SUPPORT VECTOR MACHINE



Overall test score of 65%

CONCLUSION

- Type 2 (Lodgepole Pine) is the most common type of trees in Roosevelt National Forest
- Douglas-fir has a normal distribution in terms of aspects, while Cottonwood/Willow has a left-skewed distribution where a significant number of outliers in aspects, ranging from 270-350.
- Krummholz tends to grow at a higher height, while Cottonwood/Willow tends to grow at lower height.
- Ponderosa Pine tends to grow at steeper soil, while Spruce/Fir tends to grow at flatter soil.



MODEL EVALUATION

*Overall speaking,
Random Forest Classifier does the best
job in classifying the tree types, while
Support Vector Model works perfectly in
distinguishing tree type 5, 6 and 7.*





THANKS FOR LISTENING

Q & A