# Predicting Forest Covers and Data Visualization

Team Juniper Executive Summary

Forests cover one-third of the world's surface, providing essential resources that we cannot live without in our regular lives, ranging from timber production to carbon storage. Furthermore, forests help sustain a large diversity of animals and insects. However, more forest ecosystems are now facing risks of irreversible shifts due to changing climates as represented by more severe wildfires and droughts. For these reasons, accurately assessing forest cover types is crucial to understanding the current state of forests today rather than relying on outdated data.

## Our product

The Juniper Team of the Erdős Institute has utilized machine learning algorithms to predict the dominant forest cover types from ecological and geological information about 30m x 30m patches in the Roosevelt National Forest of northern Colorado. In doing so, our project addresses two primary goals:

1. Generating an optimal algorithm to predict the cover type based on US Forest Service Information System data.
2. Visualizing the "shape" of the dataset and finding the most essential variables in clustering. We wished to communicate relationships between features to a wide audience of users.

We have evaluated the success of our models using the overall accuracies in predicting each dominant forest cover type. After testing six different algorithms, we found our XGBoost algorithm with parameter tunings had the highest accuracy, correctly predicting 88% of our data points.
We've also created images that clearly show that elevation is the feature that distinguishes clusters in our data set the most. These images highlight the most important type of data to collect in regard to cover type identification.

## Potential study case scenarios

Tree inventory data in a forest, including species, tree diameters, and heights, is essential to estimate forest productivity as well as the capacity of carbon and water storage, however, the associated work is laborious and time-consuming, and of course, it is impossible to record all the trees across a large region. Our forest cover prediction algorithm can make forest identification much easier by taking advantage of the ecological and geographic data that is already available, such as USFS Resource Information System data and MODIS satellite imaginary data from NASA. Having accurate forest cover prediction would be beneficial, for example, for park services to recognize where there is a greater density of fire-resistant tree species and a lesser chance of forest fires. Additionally, this would also help update outdated maps for military purposes without the need to send someone somewhere dangerous.

In conclusion, our machine learning algorithm gives accurate predictions for those that require an assessment of the location of cover types just from knowing certain basic features such as elevation, slope, soil types, and other features that can be found from satellite imagery rather than requiring data gatherers.