



**COSC434 – Intro to Machine Learning**  
**Course Project Proposal**

***Covertime***

**Names:** Noah Yohannes

**ID:** 100053689

**Course:** COSC434 – Intro to Machine Learning

**Submitted to:** Dr. Panagiotis Liatsis, Basit Alawode, and Ibrar Amin

**Date of Submission:** Tuesday, 31<sup>st</sup> October 2023

**Semester:** Fall 2023

## **Introduction**

Covertypes is a multiclass classification problem in machine learning that classifies a forest cover type based on values obtained from several features. Covertypes helps in developing sustainable forest management practices by predicting which species of trees are most likely to thrive in an environment. The purpose of this project is to use features that describe the state of a forest or an environment to predict the type of forest trees that are most suited to grow in such an environment.

## **Dataset**

The dataset in this project is obtained from the UC Irvine Machine learning repository. The data has 7 class labels, types of forests that are most suited to an environment described using cartographic features. 52 features are used to describe a forest, including elevation, aspect, slope, etc..... The dataset has no missing values. The feature values are not normalized so there are huge numerical variations between the features.

Some of the features of the dataset are numeric and some are categorical. The dataset has a total of 581,012 features' sets and their corresponding class label instances.

## **Approach**

Since the dataset from UCI has no missing values, the first step to solve this problem will be normalizing the features so that the pattern in the data could be observed in a smaller amount. Furthermore, it will be easier to compare relationships that exist between some features.

Then the features will be heavily visualized and analyzed. Any outlier or anomaly values observed will be either removed or replaced with an appropriate amount.

After preprocessing the data several machine learning algorithms to model the data will be explored. From those algorithms, the most appropriate one will be selected.

## **Expected Outcome**

By the end of this project a machine learning model that captures the data provided with high accuracy will be developed. This model will be able to confidently label class label of a new environment with unseen feature values.