Team Juniper

Erdős Institute Bootcamp Fall 2022

Predicting Forest Cover Types and Visualizing Data

Moeka Ono
Chris Chia
Elizabeth Campolongo (Mentor)

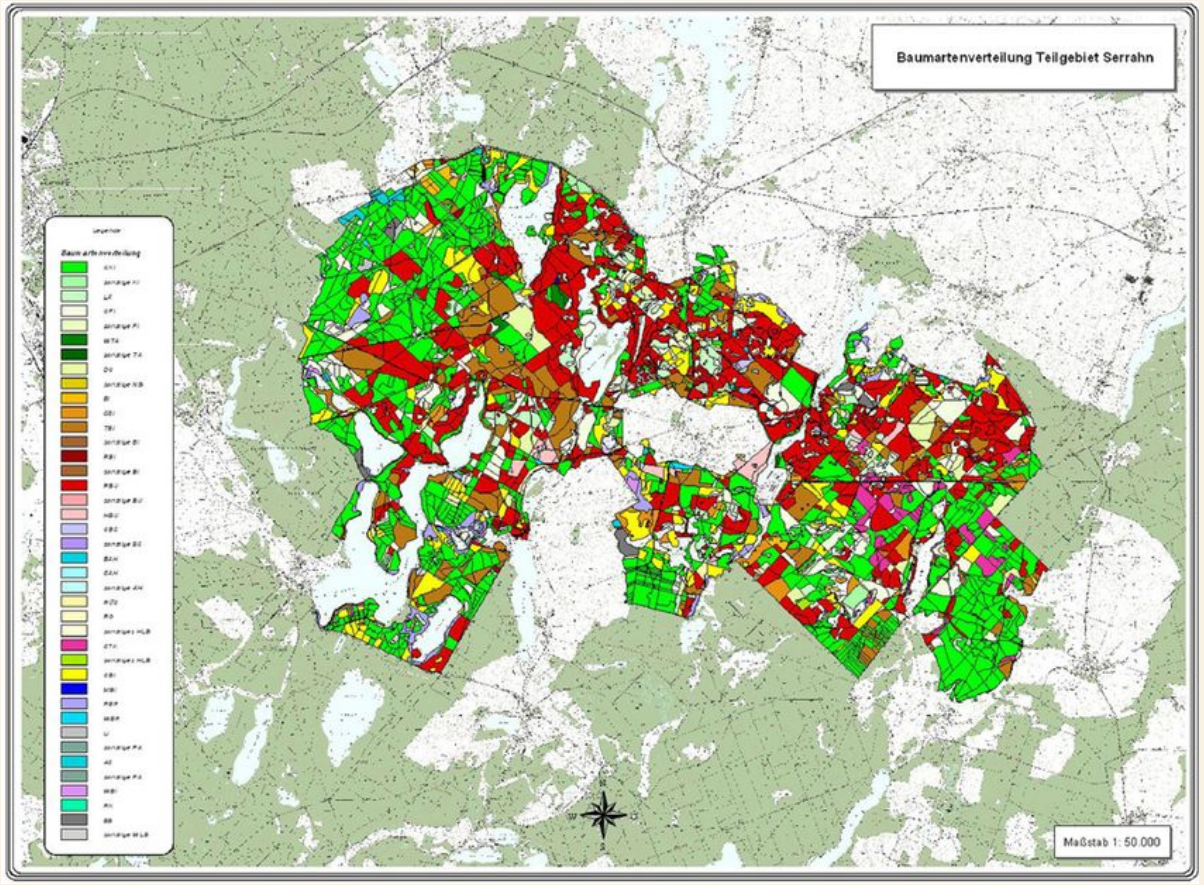
https://github.com/Erdos-Team-Juniper/forest_cover_type

The Erdős Institute

SLIDESMANIA.COM

Premise

Map of current dominating tree species in the Serrahn part of the Müritz National Park



Maps of forests are necessary!

- To know current resources
- To know locations of fire-intolerant and fire tolerant tree species
- To use for military purposes

Source: Lupp, Gerd & Konold, Werner & Bastian, Olaf. (2013). Landscape management and landscape changes towards more naturalness and wilderness

SLIDESMANIA.COM

Situation

- However, forests change over time!
  - Severe wildfires and drought due to climate change
  - Light and nutrient availabilities

- Either now or later, maps of forests will be outdated. If you want to update them…
  - Collecting new data can be expensive!
  - Collecting new data can be dangerous!

SLIDESMANIA.COM

Situation

- However, forests change over time!
  - Severe wildfires and drought due to climate change
  - Light and nutrient availabilities

- Either now or later, maps of forests will be outdated. If you want to update them…
  - Collecting new data can be expensive!
  - Collecting new data can be dangerous!

Instead, can we predict tree covers using easily accessible data?

Can we identify the most important kind of data to collect to predict tree covers?

Our data set

- 15,120 total observations. Each observation is of a 30m x 30m square in the Roosevelt National Forest of northern Colorado.

- Each 30m x 30m square has 12 descriptive features.

Elevation	Aspect	Slope	Horizontal	Vertical_D	Horizontal	Hillshade_	Hillshade_	Hillshade_	Horizontal	Wildernes	Wildernes
2596	51	3	258	0	510	221	232	148	6279	1	0
2590	56	2	212	-6	390	220	235	151	6225	1	0
2804	139	9	268	65	3180	234	238	135	6121	1	0
2785	155	18	242	118	3090	238	238	122	6211	1	0
2595	45	2	153	-1	391	220	234	150	6172	1	0
2579	132	6	300	-15	67	230	237	140	6031	1	0
2606	45	7	270	5	633	222	225	138	6256	1	0
2605	49	4	234	7	573	222	230	144	6228	1	0
2617	45	9	240	56	666	223	221	133	6244	1	0
2612	59	10	247	11	636	228	219	124	6230	1	0
2612	201	4	180	51	735	218	243	161	6222	1	0
2886	151	11	371	26	5253	234	240	136	4051	1	0

■ ■ ■

Sample of data set

SLIDESMANIA.COM

Our data set

- 15,120 total observations. Each observation is of a 30m x 30m square in the Roosevelt National Forest of northern Colorado.

- Each 30m x 30m square has 12 descriptive features.

- Each 30m x 30m square has a *DOMINANT COVER TYPE*: the species of tree that appears most often within the square.
  - 7 different cover types: Spruce, Lodgepole Pine, Aspen, etc.
  - Well-balanced data set

SLIDESMANIA.COM

Our data set

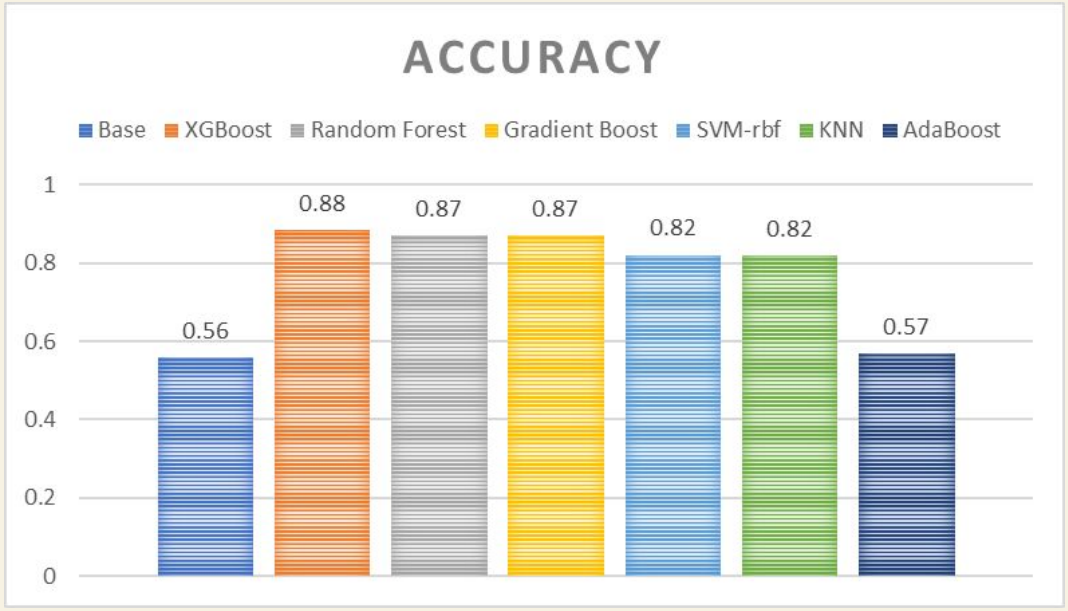
- 15,120 total observations. Each observation is of a 30m x 30m square in the Roosevelt National Forest of northern Colorado.
- Each 30m x 30m square has 12 descriptive features.
- Each 30m x 30m square has a *DOMINANT COVER TYPE*: the species of tree that appears most often within the square.
  - 7 different cover types: Spruce, Lodgepole Pine, Aspen, etc.
  - Well-balanced data set

Can we predict the dominant cover type from just these features?

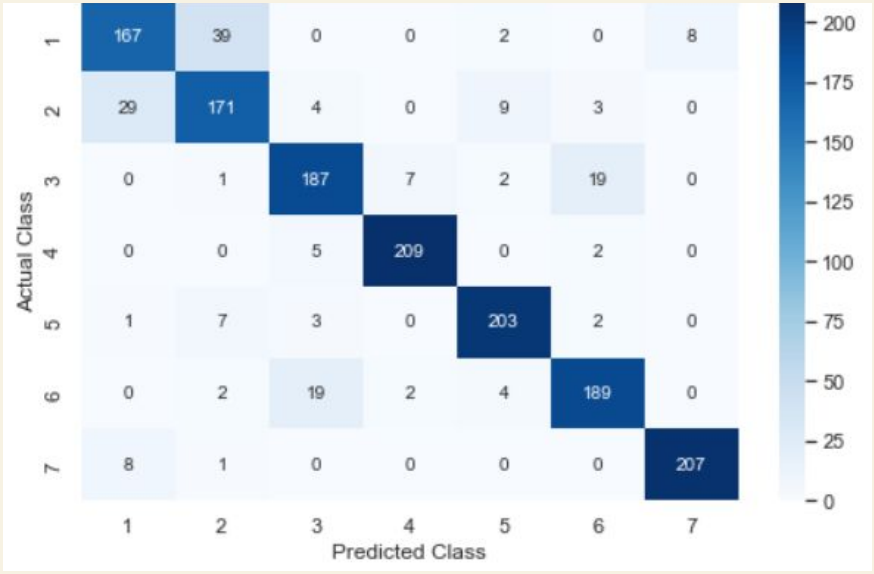
SLIDESMANIA.COM

Modelling & Results

- Used 90% of the original data for training and the rest for validation
- **XGBoosting** gave us the most accurate predictions with **88% accuracy** in predicting cover types.
- Compared with the base model (multi-logistic regression), our final model improved the accuracy by **22%.**



XGBoost



Parameter tuning:
max_depth = 8
n_estimators = 500
learning rate = 1

SLIDESMANIA.COM

Visualizing data with Topology

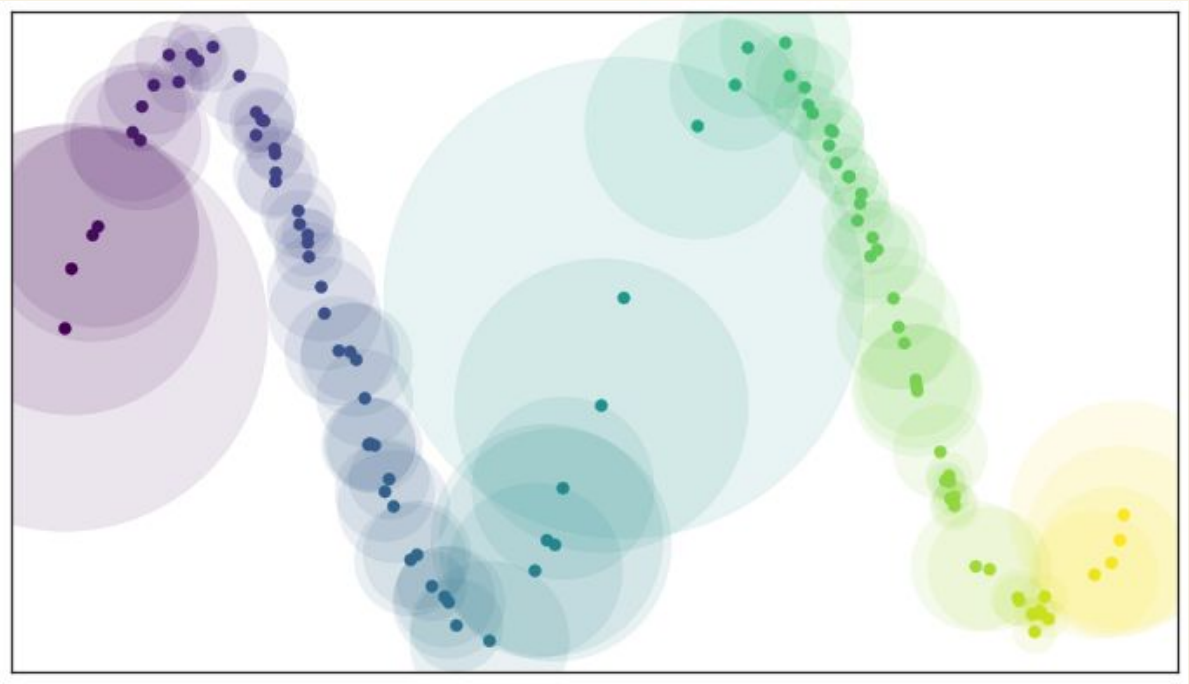
What feature is most essential in classifying tree covers?

Answered with topology! Topology is used to describe the features of a shape.

First: does our dataset have a shape? Can we describe it?

Used topology based algorithm called UMAP Algorithm

- Dimension-reduction
- Clusters points based on how CLOSE they are



Part of UMAP Algorithm

SLIDESMANIA.COM

Visualizing data with Topology

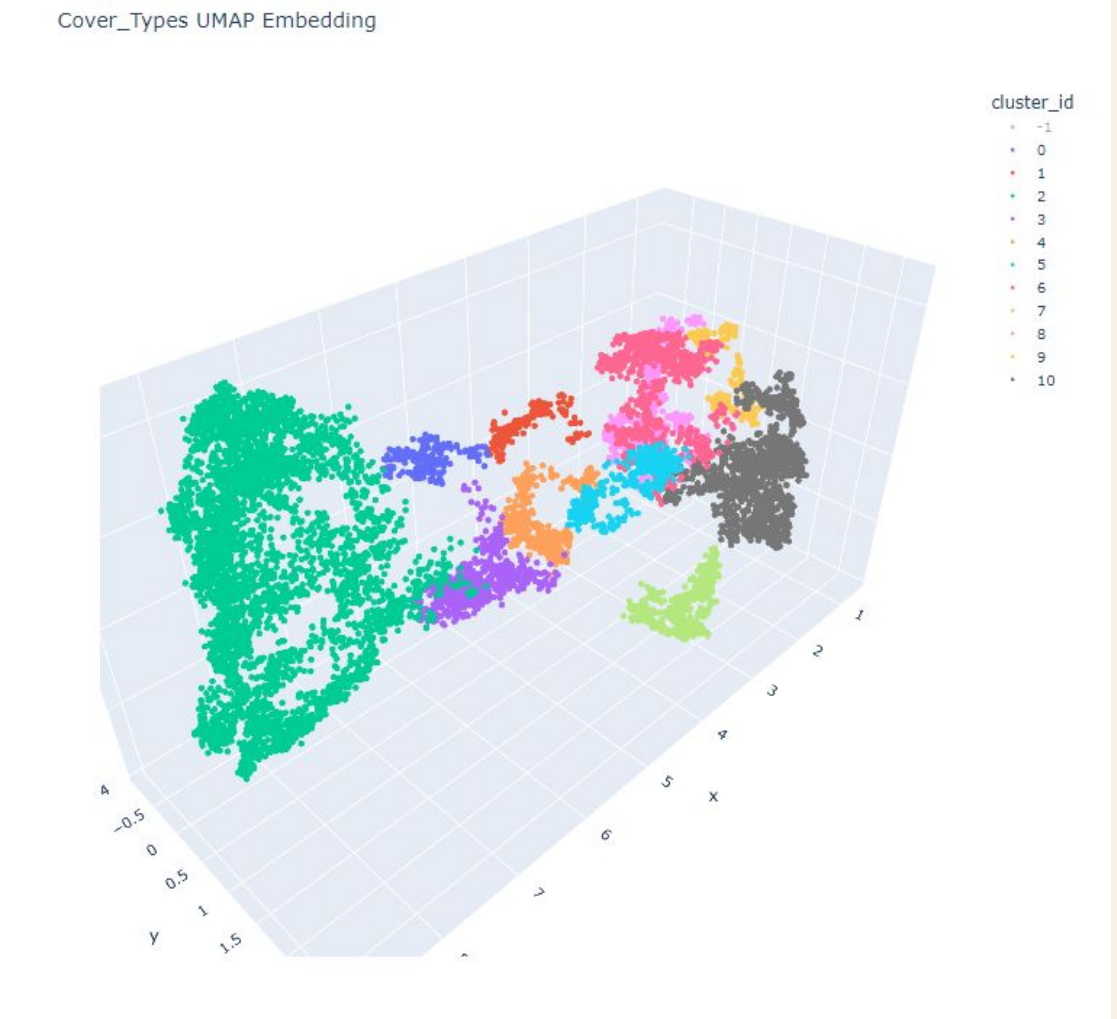
What feature is most essential in classifying cover types?

Answered with topology! Topology is used to describe the features of a shape.

First: does our dataset have a shape? Can we describe it?

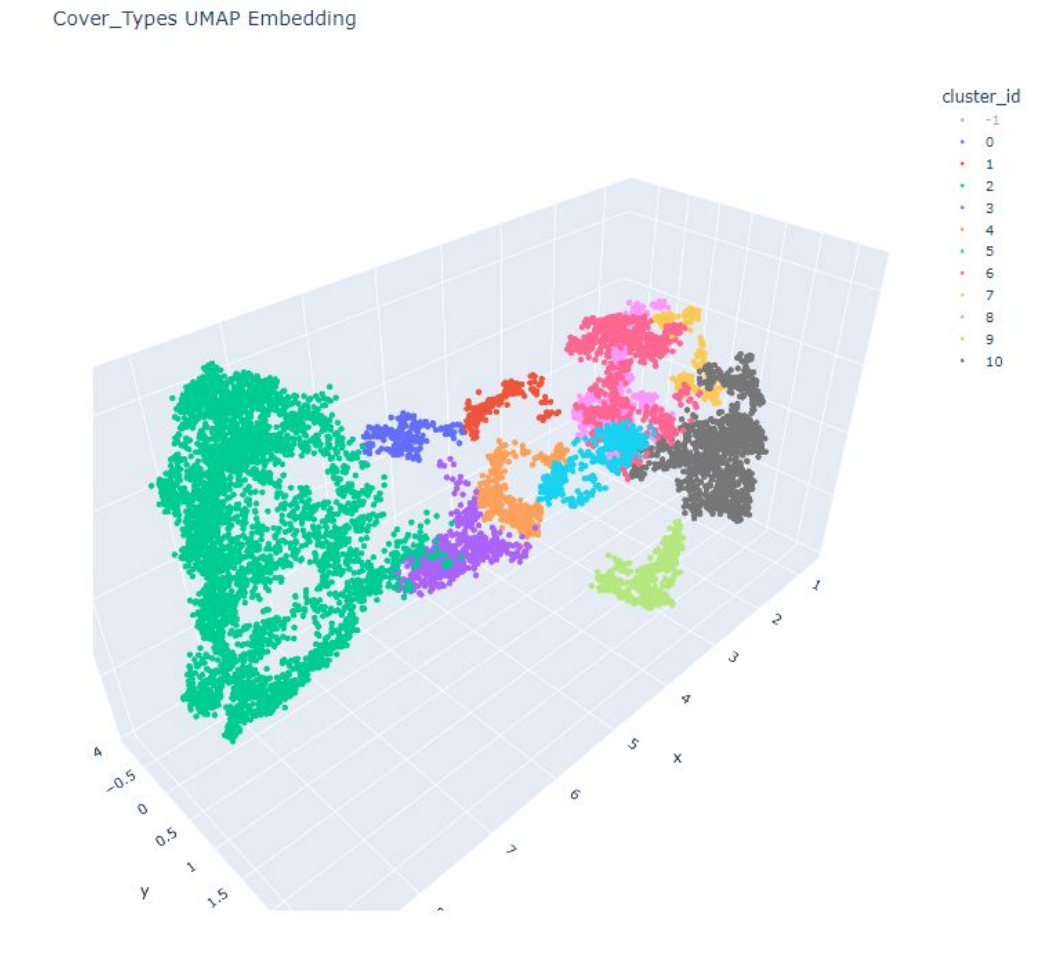
We can see our clusters are distributed along the x-axis.

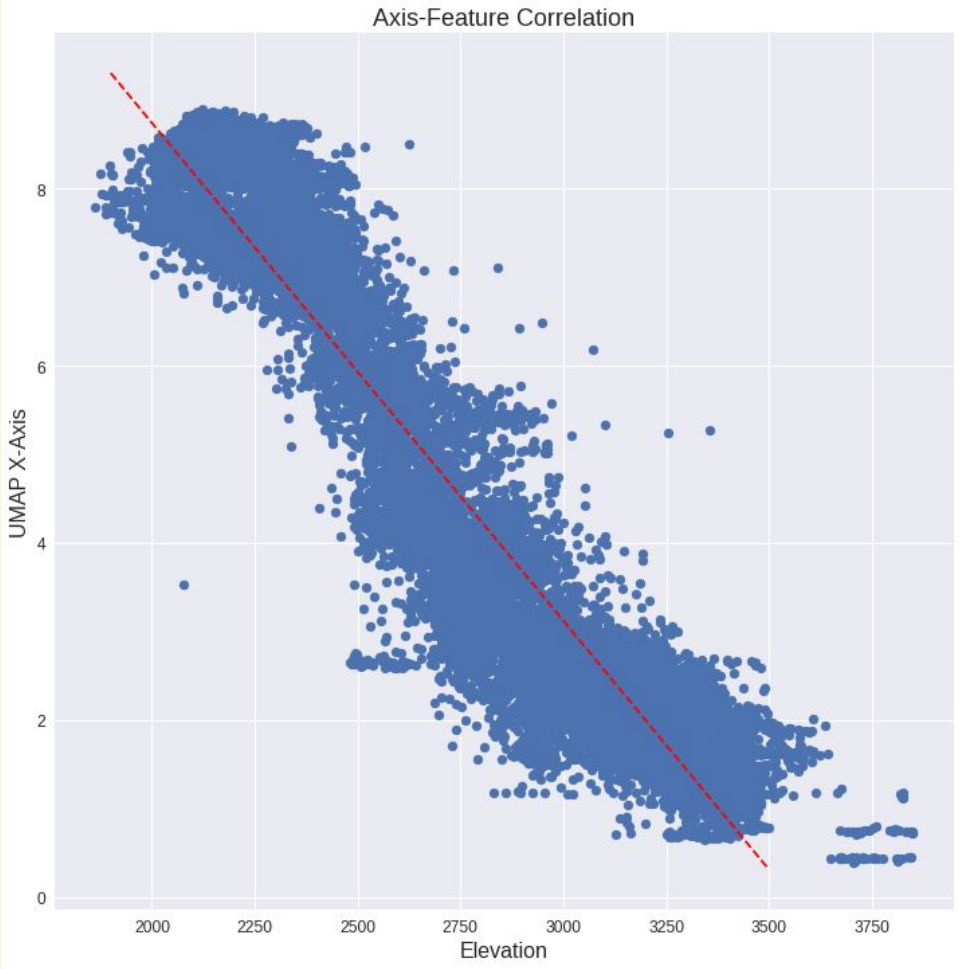
Does the x-axis relate to any of our features?



SLIDESMANIA.COM

Visualizing data with Topology





Elevation is strongly correlated with the x-axis in the shape of our dataset!

SLIDESMANIA.COM

Future Directions

- Implement time: can we predict how cover types change over time by comparing with data sets from the past?
- Better accuracy in distinguishing Spruce/Fir and Lodgepole Pine cover types
  - Our model struggled the most with these two types
- More parameter tuning on topological model to separate into 7 clusters for each cover type.

SLIDESMANIA.COM

THANK YOU!

Special thanks to Elizabeth for her helpful feedback!

SLIDESMANIA.COM