

Primera Entrega de Proyecto

Carolina Morales Galeano

Facultad de Ingeniería, Universidad de Antioquía

Inteligencia Artificial para Ingenieros

Raúl Ramos Pollan

Marzo 11,2023

1. Planteamiento del problema

En este proyecto se predecirá el tipo de cubierta forestal (el tipo predominante de cubierta arbórea) a partir de variables estrictamente cartográficas de cuatro áreas silvestres ubicadas en el Bosque Roosevelt del norte de Colorado.

2. Dataset

El dataset a utilizar proviene de una competencia de kaggle en la cual se proporciona información de cuatro áreas silvestres ubicadas en el Bosque Nacional Roosevelt del norte de Colorado. Cada observación es un parche de 30m x 30m. Se le pide que prediga una clasificación entera para el tipo de cubierta forestal. Los siete tipos de cubierta forestal que se predicarán son:

- Picea/abeto
- Pino Lodgepole
- Pino Ponderosa
- Álamo/Sauce
- Álamo temblón
- Abeto de Douglas
- Krummholz

El conjunto de entrenamiento (15120 observaciones) contiene tanto entidades como Cover_Type. El conjunto de prueba contiene solo las características. Debe predecir Cover_Type para cada fila en el conjunto de prueba (565892 observaciones).

Campos de información

- ***Elevación*** : elevación en metros
- ***Aspecto*** : aspecto en grados de acimut
- ***Pendiente*** : pendiente en grados
- ***Distancia_horizontal*** a la hidrología: distancia horizontal a las características de agua superficial más cercanas
- ***Distancia_vertical*** a la hidrología: distancia vertical a las características de agua superficial más cercanas
- ***Distancia_horizontal*** a las carreteras: distancia horizontal a la carretera más cercana
- ***Sombreado_9*** a. índice a las 9 a. m., solsticio de verano

- **Hillshade_Noon** (índice de 0 a 255) - Índice de sombreado al mediodía, solsticio de verano
- **Hillshade_3pm** (índice de 0 a 255) - Índice de sombreado a las 3 p. m., solsticio de verano
- **Horizontal_Distance_To_Fire_Points**- Dist Horz a los puntos de ignición de incendios forestales más cercanos
- **Wilderness_Area** (4 columnas binarias, 0 = ausencia o 1 = presencia) - Designación de área silvestre
- **Soil_Type** (40 columnas binarias, 0 = ausencia o 1 = presencia) - Designación de tipo de suelo
- **Cover_Type** (7 tipos, números enteros 1 a 7) - Designación del tipo de cubierta forestal

Las áreas silvestres son:

- Área silvestre de Rawah
- Área silvestre de Neota
- Área silvestre de Comanche Peak
- Área silvestre de Cache la Poudre

Los tipos de suelo son:

- Familia Catedral - Conjunto de afloramientos rocosos, extremadamente pedregosos.
- Vanet - Complejo de familias Ratake, muy pedregoso.
- Haploborolis - Complejo de afloramientos rocosos, escombros.
- Familia Ratake - Complejo de afloramientos rocosos, escombros.
- Familia Vanet - Complejo complejo de afloramientos rocosos, escombros. Vanet
- Familias Wetmore - Complejo de afloramientos rocosos, pedregosos.
- Familia gótica.
- Supervisor - Complejo de familias Limber.
- Familia Troutville, muy pedregosa.
- Bullwark - Familias Catamount - Complejo de afloramientos rocosos, escombros.
- Bullwark - Familias Catamount - Complejo de terreno rocoso, escombros.
- Familia Legault - Complejo rocoso, pedregoso.
- Familia Catamount - Rock land - Complejo familiar Bullwark, escombros.
- Complejo Pachic Argiborolis - Aquolis.

- no especificados en el USFS Soil and ELU Survey.
- Cryaquolis - complejo Cryoborolis.
- Familia Gateview - Complejo Cryaquolis.
- Familia Rogert, muy pedregoso.
- Typic Cryaquolis - Borohemists complejo.
- Typic Cryaquepts - Complejo Typic Cryaquolls.
- Typic Cryaquolls - Familia Leighcan, hasta complejo de sustrato.
- Familia Leighcan, hasta sustrato, extremadamente pedregoso.
- Familia Leighcan, hasta sustrato - complejo Typic Cryaquolls.
- Familia Leighcan, extremadamente pedregosa.
- Familia Leighcan, cálido, extremadamente pedregoso.
- Granile - Conjunto de familias Catamount, muy pedregoso.
- Familia Leighcan, cálida - Complejo de afloramientos rocosos, extremadamente pedregosos.
- Familia Leighcan - Complejo de afloramientos rocosos, extremadamente pedregosos.
- Como - Complejo de familias Legault, extremadamente pedregoso.
- Familia Como - Terreno rocoso - Complejo familiar Legault, extremadamente pedregoso.
- Leighcan - Complejo de familias Catamount, extremadamente pedregoso.
- Familia Catamount - Afloramiento rocoso - Complejo de la familia Leighcan, extremadamente pedregoso.
- Leighcan - Familias Catamount - Complejo de afloramientos rocosos, extremadamente pedregosos.
- Cryorthents - Complejo de terreno rocoso, extremadamente pedregoso.
- Cryumbrepts - Afloramiento rocoso - Complejo Cryaquepts.
- Familia Bross - Tierra de rocas - Complejo Cryumbrepts, extremadamente pedregoso.
- Afloramiento rocoso - Cryumbrepts - Complejo Cryorthents, extremadamente pedregoso.
- Leighcan - Familias Moran - Complejo Cryaquolls, extremadamente pedregoso.
- Familia Moran - Cryorthents - Complejo de la familia Leighcan, extremadamente pedregoso.
- Familia Moran - Cryorthents - Complejo de terreno rocoso, extremadamente pedregoso.

3. Métricas

La métrica de evaluación principal para el modelo será **Accuracy**, la cual representa el porcentaje total de valores correctamente clasificados, tanto positivos como negativos.

Por otra parte, en cuanto a la métrica de negocio, se tiene interés en que las predicciones sean lo suficientemente confiables para saber la cantidad de CO₂ absorbido por la cobertura forestal (ya que depende de la especie que compone la misma en su mayoría), y de esta forma cuantificar el impacto ambiental de una deforestación y la descarbonización del ambiente en una zona determinada.

4. Desempeño

Si la Tasa Interna de Retorno del proyecto anual, no es mayor o igual al 28.5%, no sería rentable poner el modelo en producción, ya que el coste de implementación y mantenimiento, no cubriría los costes del mismo.

5. Bibliografías

Forest Cover Type Prediction / Kaggle. (s. f.).<https://www.kaggle.com/competitions/forest-cover-type-prediction/overview>

