**Overview**

*Another weekend and another exciting hackathon, and this time with an open dataset. Yes, you heard it right !*

The dataset has been taken from [UCI](https://archive.ics.uci.edu/ml/index.php), but to keep the spirit of competition right, we have added some noise in the labels. In this hackathon, we challenge all Machinehackers to predict the forest cover types (the predominant kind of tree cover) from strictly cartographic variables (as opposed to remotely sensed data).

The actual forest cover type for a given 30 x 30-meter cell was determined from US Forest Service (USFS) Region to Resource Information System data. Independent variables were then derived from the data obtained from the US Geological Survey and USFS.

The data is in raw form (not scaled) and contains binary columns of data for qualitative independent variables such as wilderness areas and soil type (one-hot-encoded).

This study area includes four wilderness areas located in the Roosevelt National Forest of northern Colorado.

**Dataset Description:**

* **train.csv - 29050 rows x 55 columns**
* **test.csv - 551962 rows x 54 columns**
* **sample\_submission.csv - Accepted submission format**

**Attribute Information:**

* **Elevation** - Elevation in meters
* **Aspect** - Aspect in degrees
* **Slope** - Slope in degrees
* **Horizontal\_Distance\_To\_Hydrology** - Horz Dist to nearest surface water features
* **Vertical\_Distance\_To\_Hydrology** - Vert Dist to nearest surface water features
* **Horizontal\_Distance\_To\_Roadways** - Horz Dist to the nearest roadway
* **Hillshade\_9am** (0 to 255 index) - Hillshade index at 9am, summer solstice
* **Hillshade\_Noon** (0 to 255 index) - Hillshade index at noon, summer solstice
* **Hillshade\_3pm** (0 to 255 index) - Hillshade index at 3pm, summer solstice
* **Horizontal\_Distance\_To\_Fire\_Points** - Horz Dist to nearest wildfire ignition points
* **Wilderness\_Area** (4 binary columns, 0 = absence or 1 = presence) - Wilderness area designation
* **Soil\_Type** (40 binary columns, 0 = absence or 1 = presence) - Soil Type designation
* **Cover\_Type** (7 types, integers 1 to 7) - Forest Cover Type designation
* The wilderness areas are:
  + **1** - Rawah Wilderness Area
  + **2** - Neota Wilderness Area
  + **3** - Comanche Peak Wilderness Area
  + **4** - Cache la Poudre Wilderness Area
* The soil types are:
  + **1 Cathedral family** - Rock outcrop complex, extremely stony.
  + **2 Vanet** - Ratake families complex, very stony.
  + **3 Haploborolis** - Rock outcrop complex, rubble.
  + **4 Ratake family** - Rock outcrop complex, rubble.
  + **5 Vanet family** - Rock outcrop complex, rubble.
  + **6 Vanet** - Wetmore families - Rock outcrop complex, stony.
  + **7 Gothic family**.
  + **8 Supervisor** - Limber families complex.
  + **9 Troutville family**, very stony.
  + **10 Bullwark** - Catamount families - Rock outcrop complex, rubble.
  + **11 Bullwark** - Catamount families - Rock land complex, rubble.
  + **12 Legault family** - Rock land complex, stony.
  + **13 Catamount family**- Rock land - Bulwark family complex, rubble.
  + **14 Pachic Argiborolis** - Aquolis complex.
  + **15** unspecified in the USFS Soil and ELU Survey.
  + **16 Cryaquolis** - Cryoborolis complex.
  + **17 Gateview family** - Cryaquolis complex.
  + **18 Rogert family,** very stony.
  + **19 Typic Cryaquolis** - Borohemists complex.
  + **20 Typic Cryaquepts** - Typic Cryaquolls complex.
  + **21 Typic Cryaquolls** - Leighcan family, till substratum complex.
  + **22 Leighcan family,** till substratum, extremely boulder.
  + **23 Leighcan family**, till substratum - Typic Cryaquolls complex.
  + **24 Leighcan family**, extremely stony.
  + **25 Leighcan family**, warm, extremely stony.
  + **26 Granile** - Catamount families complex, very stony.
  + **27 Leighcan family**, warm - Rock outcrop complex, extremely stony.
  + **28 Leighcan family**- Rock outcrop complex, extremely stony.
  + **29 Como**- Legault families complex, extremely stony.
  + **30 Como family** - Rock land - Legault family complex, extremely stony.
  + **31 Leighcan**- Catamount families complex, extremely stony.
  + **32 Catamount family** - Rock outcrop - Leighcan family complex, extremely stony.
  + **33 Leighcan** - Catamount families - Rock outcrop complex, extremely stony.
  + **34 Cryorthents** - Rock land complex, extremely stony.
  + **35 Cryumbrepts** - Rock outcrop - Cryaquepts complex.
  + 3**6 Bross family** - Rock land - Cryumbrepts complex, extremely stony.
  + **37 Rock outcrop** - Cryumbrepts - Cryorthents complex, extremely stony.
  + **38 Leighcan** - Moran families - Cryaquolls complex, extremely stony.
  + **39 Moran family** - Cryorthents - Leighcan family complex, extremely stony.
  + **40 Moran family** - Cryorthents - Rock land complex, extremely stony.

**ACKNOWLEDGEMENTS:**

*MachineHack is hosting this hackathon for the data science community worldwide just for fun, practice and learning.*

*This dataset was provided by Jock A. Blackard and Colorado State University.*

*We also thank the UCI machine learning repository for*[*hosting the dataset*](https://archive.ics.uci.edu/ml/datasets/Covertype)*.*

***Original Owners of Database:***

*Remote Sensing and GIS Program*  
*Department of Forest Sciences*  
*College of Natural Resources*  
*Colorado State University*  
*Fort Collins, CO 80523*  
*(contact Jock A. Blackard, jblackard****@****fs.fed.us or Dr. Denis J. Dean, denis.dean****@****utdallas.edu)*