

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

[Center for Machine Learning and Intelligent Systems](#)
 

☒ Repository
 ☐ Web
 

[View ALL Data Sets](#)

## Covertypes Data Set

Download: [Data Folder](#), [Data Set Description](#)

**Abstract:** Forest CoverType dataset



<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	581012	<b>Area:</b>	Life
<b>Attribute Characteristics:</b>	Categorical, Integer	<b>Number of Attributes:</b>	54	<b>Date Donated</b>	1998-08-01
<b>Associated Tasks:</b>	Classification	<b>Missing Values?</b>	No	<b>Number of Web Hits:</b>	234506

### Source:

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](mailto:jblackard@fs.fed.us) or Dr. Denis J. Dean, [denis.dean '@' utdallas.edu](mailto:denis.dean@utdallas.edu))

Donors of database:

1. Jock A. Blackard ([jblackard '@' fs.fed.us](mailto:jblackard@fs.fed.us))  
GIS Coordinator  
USFS - Forest Inventory & Analysis  
Rocky Mountain Research Station  
507 25th Street  
Ogden, UT 84401

2. Dr. Denis J. Dean ([denis.dean '@' utdallas.edu](mailto:denis.dean@utdallas.edu))  
Professor  
Program in Geography and Geospatial Sciences  
School of Economic, Political and Policy Sciences  
800 West Campbell Rd  
Richardson, TX 75080-3021

3. Dr. Charles W. Anderson ([anderson '@' cs.colostate.edu](mailto:anderson@cs.colostate.edu))  
Associate Professor  
Department of Computer Science

Colorado State University  
Fort Collins, CO 80523 USA

## Data Set Information:

Predicting forest cover type from cartographic variables only (no remotely sensed data). The actual forest cover type for a given observation (30 x 30 meter cell) was determined from US Forest Service (USFS) Region 2 Resource Information System (RIS) data. Independent variables were derived from data originally obtained from US Geological Survey (USGS) and USFS data. Data is in raw form (not scaled) and contains binary (0 or 1) columns of data for qualitative independent variables (wilderness areas and soil types).

This study area includes four wilderness areas located in the Roosevelt National Forest of northern Colorado. These areas represent forests with minimal human-caused disturbances, so that existing forest cover types are more a result of ecological processes rather than forest management practices.

Some background information for these four wilderness areas: Neota (area 2) probably has the highest mean elevational value of the 4 wilderness areas. Rawah (area 1) and Comanche Peak (area 3) would have a lower mean elevational value, while Cache la Poudre (area 4) would have the lowest mean elevational value.

As for primary major tree species in these areas, Neota would have spruce/fir (type 1), while Rawah and Comanche Peak would probably have lodgepole pine (type 2) as their primary species, followed by spruce/fir and aspen (type 5). Cache la Poudre would tend to have Ponderosa pine (type 3), Douglas-fir (type 6), and cottonwood/willow (type 4).

The Rawah and Comanche Peak areas would tend to be more typical of the overall dataset than either the Neota or Cache la Poudre, due to their assortment of tree species and range of predictive variable values (elevation, etc.) Cache la Poudre would probably be more unique than the others, due to its relatively low elevation range and species composition.

## Attribute Information:

Given is the attribute name, attribute type, the measurement unit and a brief description. The forest cover type is the classification problem. The order of this listing corresponds to the order of numerals along the rows of the database.

Name / Data Type / Measurement / Description

Elevation / quantitative / meters / Elevation in meters  
 Aspect / quantitative / azimuth / Aspect in degrees azimuth  
 Slope / quantitative / degrees / Slope in degrees  
 Horizontal\_Distance\_To\_Hydrology / quantitative / meters / Horz Dist to nearest surface water features  
 Vertical\_Distance\_To\_Hydrology / quantitative / meters / Vert Dist to nearest surface water features  
 Horizontal\_Distance\_To\_Roadways / quantitative / meters / Horz Dist to nearest roadway  
 Hillshade\_9am / quantitative / 0 to 255 index / Hillshade index at 9am, summer solstice  
 Hillshade\_Noon / quantitative / 0 to 255 index / Hillshade index at noon, summer solstice  
 Hillshade\_3pm / quantitative / 0 to 255 index / Hillshade index at 3pm, summer solstice  
 Horizontal\_Distance\_To\_Fire\_Points / quantitative / meters / Horz Dist to nearest wildfire ignition points  
 Wilderness\_Area (4 binary columns) / qualitative / 0 (absence) or 1 (presence) / Wilderness area designation  
 Soil\_Type (40 binary columns) / qualitative / 0 (absence) or 1 (presence) / Soil Type designation  
 Cover\_Type (7 types) / integer / 1 to 7 / Forest Cover Type designation

## Relevant Papers:

Blackard, Jock A. and Denis J. Dean. 2000. "Comparative Accuracies of Artificial Neural Networks and Discriminant Analysis in Predicting Forest Cover Types from Cartographic Variables." Computers and Electronics in Agriculture 24(3):131-151.  
[\[Web Link\]](#)

Blackard, Jock A. and Denis J. Dean. 1998. "Comparative Accuracies of Neural Networks and Discriminant Analysis in Predicting Forest Cover Types from Cartographic Variables." Second Southern Forestry GIS Conference. University of Georgia. Athens, GA. Pages 189-199.

Blackard, Jock A. 1998. "Comparison of Neural Networks and Discriminant Analysis in Predicting Forest Cover Types." Ph.D. dissertation. Department of Forest Sciences. Colorado State University. Fort Collins, Colorado. 165 pages.

## Papers That Cite This Data Set<sup>1</sup>:



Joao Gama and Ricardo Rocha and Pedro Medas. [Accurate decision trees for mining high-speed data streams](#). KDD. 2003. [\[View Context\]](#).

Nikunj C. Oza and Stuart J. Russell. [Experimental comparisons of online and batch versions of bagging and boosting](#). KDD. 2001. [\[View Context\]](#).

Johannes Furnkranz. [Round Robin Rule Learning](#). Austrian Research Institute for Artificial Intelligence. [\[View Context\]](#).

Zoran Obradovic and Slobodan Vucetic. [Challenges in Scientific Data Mining: Heterogeneous, Biased, and Large Samples](#). Center for Information Science and Technology Temple University. [\[View Context\]](#).

Arto Klami and Samuel Kaski and Ty n ohjaaja and Janne Sinkkonen. [HELSINKI UNIVERSITY OF TECHNOLOGY Department of Engineering Physics and Mathematics Arto Klami Regularized Discriminative Clustering](#). Regularized Discriminative Clustering. [\[View Context\]](#).

Chris Giannella and Bassem Sayrafi. [An Information Theoretic Histogram for Single Dimensional Selectivity Estimation](#). Department of Computer Science, Indiana University Bloomington. [\[View Context\]](#).

## Citation Request:

Reuse of this database is unlimited with retention of copyright notice for Jock A. Blackard and Colorado State University.

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with [Rexa.info](#)

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#). || [Donation Policy](#). || [Contact](#) || [CML](#)