Attributes :

[1] "Elevation"

[2] "Aspect"

[3] "Slope"

[4] "Horizontal\_Distance\_To\_Hydrology"

[5] "Vertical\_Distance\_To\_Hydrology"

[6] "Horizontal\_Distance\_To\_Roadways"

[7] "Hillshade\_9am"

[8] "Hillshade\_Noon"

[9] "Hillshade\_3pm"

[10] "Horizontal\_Distance\_To\_Fire\_Points"

[11] "Area\_Of\_Wilderness"

[12] "Soil\_Type"

[13] "Cover\_Type"

**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, rubbly.

4 Ratake family - Rock outcrop complex, rubbly.

5 Vanet family - Rock outcrop complex complex, rubbly.

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, rubbly.

11 Bullwark - Catamount families - Rock land complex, rubbly.

12 Legault family - Rock land complex, stony.

13 Catamount family - Rock land - Bullwark family complex, rubbly.

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 bouldery.

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.

36 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.

**Transformation Description:**

**Transform1.csv** : Converts the raw training data into reduced format, merge asymmetric binary attributes of soil type and area of wilderness into single row

**test\_transform1.csv** : For test data

**Code** : forestDataLib.R and transformation.R

**Transform2.csv** : Converts the transform1 data into standardized form(gaussian distribution),

with mean 0, standard deviation =1

**test\_transform2.csv** : For test data

**Code** : forestDataLib.R and transformation.R

**Transform3.csv** : Converts the transform2 nominal attributes into nominal labels from integers.

Eg: soil-type S1 … S40

area\_of\_wilderness A1 … A4

cover\_type C1.. C7

It will be used mostly for classification as many classifiers do not identify integers as nominal data.

**test\_transform3.csv** : For test data

**Code:** forestDataLib.R and transformation.R

**Transform4.csv** Converts the transform1 data into min-max normalized form with transformation of numerical nominal data to categorical , where the min-max range is taken from test\_transform1 data.

**test\_transform4.csv :** For test data

**Code:** forestDataLib.R and transformation.R

**Transform5.csv** Converts the transform1 data into min-max normalized form, without any nominal transformation where the min-max range is taken from test\_transform1 data.

**test\_transform5.csv :** For test data

**Code:** forestDataLib.R and transformation.R

**Transform6.csv** Training data doesn’t contain instances for soil type 7 and soil type 15. Transform6 contains some pure instances for these soil type added from test data based on the results of various classifiers.

**test\_transform4.csv :** For test data

Enhance1.csv : Combining pure prediction results of different classifiers from test data to train data. Around 47% of the test\_data has pure prediction results.

Enhance2.csv : Taking 40,000 pure predictions from test to train set.

Extract1.csv : Extracted few important features from transform5.csv

Transform7.csv : Keeping the original data with binary attributes and adding few missing soil type data from test data. Normalized data. Cover type is kept as nominal.

Test Transform7.csv : Same as transform7 but for test data

Transform8.csv : Same as transform7 but numeric data as covertype