**Elastic Multi Stage Decision Methodology**

The work is an outgrowth of Dr. Datta and Dr. Mengel’s investigation to find rules with better accuracy for the infrequent group. In the investigation, different decision tree algorithms, such as C4.5, CART, Recursive Partition, and J48graft, were applied on different datasets. The code that we generated in R language is based on recursive partitioning to find metrics at each level of split in a decision tree. In MSDM, the datasets are first divided into three clusters using K-Means clustering. Then each cluster is partitioned by a controlled/monitored decision tree using recursive partition to keep data set attributes from being reused and to abbreviate the number of levels in the tree when accuracy is not improved by going to another level. Each dataset partition at the end of a tree branch is analyzed further by utilizing association mining. The work mainly focused on recursive partitioning where the tree is built at each split and metrics were calculated and checking for the accuracy levels for further splitting.

**Installation -----------------**

**Software requirements:**

R-Studio

**Steps to execute:**

1. Installing the following packages

**install.packages("rpart")** #Recursive partitioning for classification,regression and survival trees

**install.packages("rpart.plot")** #Plot 'rpart' models. Extends plot.rpart()

**install.packages("partykit")** # A toolkit with infrastructure for representing, summarizing, and visualizing tree-structured regression and classification models.

**install.packages("Metrics")** # Metrics is a set of evaluation metrics that is commonly used in supervised machine learning.

**install.packages("gWidgets2")**  # The API is defined in this package.

**install.packages("gWidgets2tcltk")**

1. Loading the libraries

library(rpart)

library(rpart.plot)

library(partykit)

library(Metrics)

require(gWidgets2)

require(gWidgets2tcltk)

1. Execute the following functions :

**EMSDM\_Window <- function(EMSDM\_formula,EMSDM\_dataset)**

#function which builds the tree

**EMSDM\_Metrics <- function(fit, data\_metrics)**

#function which calculates the metrics

1. Accuracy

2. Kappa

3. Per-class Precision

4. Recall

5. F-1

**EMSDM\_node\_observations <- function(fit, i, EMSDM\_dataset)**

#function holds two datasets after splitting main dataset.

1. Run the function

#variable\_name

EMSDM\_dataset <- read.csv(file.choose(),header = TRUE)

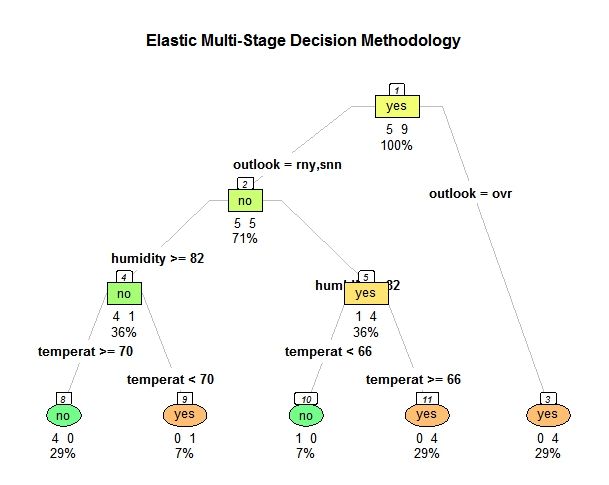
EMSDM\_formula <- EMSDM\_dataset$ variable\_name ~ .

#function call for Building Tree

EMSDM\_Window(EMSDM\_formula,EMSDM\_dataset)

**Experimental results**

**Weather dataset:**



> #function call for Building Tree

> EMSDM\_Window(EMSDM\_formula,EMSDM\_dataset)

[1] "Number of instances :14"

[1] "Correctly Classified Instances :9 64.2857142857143%"

[1] "InCorrectly Classified Instances :5 35.7142857142857%"

[1] "Comparision Matrix :"

predicted

no yes

no 0 5

yes 0 9

[1] "Accuracy :64.2857142857143%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

no NaN 0 NaN

yes 0.6428571 1 0.7826087

[1] "Kappa :0"

[1] 0

[1] "Number of instances :14"

[1] "Correctly Classified Instances :9 64.2857142857143%"

[1] "InCorrectly Classified Instances :5 35.7142857142857%"

[1] "Comparision Matrix :"

predicted

no yes

no 5 0

yes 5 4

[1] "Accuracy :64.2857142857143%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

no 0.5 1.0000000 0.6666667

yes 1.0 0.4444444 0.6153846

[1] "Kappa :0.363636363636364"

[1] 1

[1] "Number of instances :14"

[1] "Correctly Classified Instances :12 85.7142857142857%"

[1] "InCorrectly Classified Instances :2 14.2857142857143%"

[1] "Comparision Matrix :"

predicted

no yes

no 4 1

yes 1 8

[1] "Accuracy :85.7142857142857%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

no 0.8000000 0.8000000 0.8000000

yes 0.8888889 0.8888889 0.8888889

[1] "Kappa :0.688888888888889"

[1] 2

[1] "Number of instances :14"

[1] "Correctly Classified Instances :14 100%"

[1] "InCorrectly Classified Instances :0 0%"

[1] "Comparision Matrix :"

predicted

no yes

no 5 0

yes 0 9

[1] "Accuracy :100%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

no 1 1 1

yes 1 1 1

[1] "Kappa :1"

[1] 3

[1] "Number of instances :14"

[1] "Correctly Classified Instances :14 100%"

[1] "InCorrectly Classified Instances :0 0%"

[1] "Comparision Matrix :"

predicted

no yes

no 5 0

yes 0 9

[1] "Accuracy :100%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

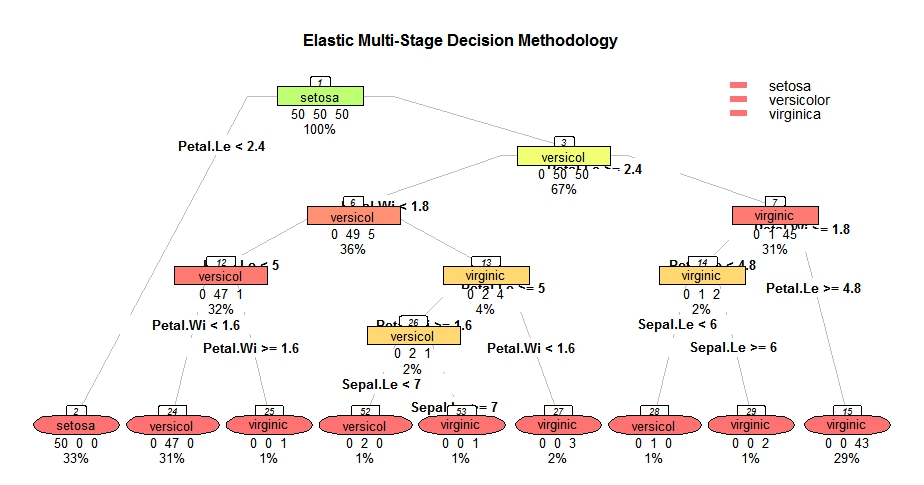
no 1 1 1

yes 1 1 1

[1] "Kappa :1"

--------------------------------------

**Iris dataset**



> #function call for Building Tree

> EMSDM\_Window(EMSDM\_formula,EMSDM\_dataset)

[1] "Number of instances :150"

[1] "Correctly Classified Instances :50 33.3333333333333%"

[1] "InCorrectly Classified Instances :100 66.6666666666667%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 50 0 0

virginica 50 0 0

[1] "Accuracy :33.3333333333333%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 0.3333333 1 0.5

versicolor NaN 0 NaN

virginica NaN 0 NaN

[1] "Kappa :0"

[1] 0

[1] "Number of instances :150"

[1] "Correctly Classified Instances :100 66.6666666666667%"

[1] "InCorrectly Classified Instances :50 33.3333333333333%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 50 0

[1] "Accuracy :66.6666666666667%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0 1 1.0000000

versicolor 0.5 1 0.6666667

virginica NaN 0 NaN

[1] "Kappa :0.5"

[1] 1

[1] "Number of instances :150"

[1] "Correctly Classified Instances :144 96%"

[1] "InCorrectly Classified Instances :6 4%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 49 1

virginica 0 5 45

[1] "Accuracy :96%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.0000000

versicolor 0.9074074 0.98 0.9423077

virginica 0.9782609 0.90 0.9375000

[1] "Kappa :0.94"

[1] 2

[1] "Number of instances :150"

[1] "Correctly Classified Instances :146 97.3333333333333%"

[1] "InCorrectly Classified Instances :4 2.66666666666667%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 47 3

virginica 0 1 49

[1] "Accuracy :97.3333333333333%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.0000000

versicolor 0.9791667 0.94 0.9591837

virginica 0.9423077 0.98 0.9607843

[1] "Kappa :0.96"

[1] 3

[1] "Number of instances :150"

[1] "Correctly Classified Instances :144 96%"

[1] "InCorrectly Classified Instances :6 4%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 49 1

virginica 0 5 45

[1] "Accuracy :96%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.0000000

versicolor 0.9074074 0.98 0.9423077

virginica 0.9782609 0.90 0.9375000

[1] "Kappa :0.94"

[1] 2

[1] "Number of instances :150"

[1] "Correctly Classified Instances :100 66.6666666666667%"

[1] "InCorrectly Classified Instances :50 33.3333333333333%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 50 0

[1] "Accuracy :66.6666666666667%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0 1 1.0000000

versicolor 0.5 1 0.6666667

virginica NaN 0 NaN

[1] "Kappa :0.5"

[1] 1

[1] "Number of instances :150"

[1] "Correctly Classified Instances :50 33.3333333333333%"

[1] "InCorrectly Classified Instances :100 66.6666666666667%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 50 0 0

virginica 50 0 0

[1] "Accuracy :33.3333333333333%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 0.3333333 1 0.5

versicolor NaN 0 NaN

virginica NaN 0 NaN

[1] "Kappa :0"

[1] 0

[1] "Number of instances :150"

[1] "Correctly Classified Instances :100 66.6666666666667%"

[1] "InCorrectly Classified Instances :50 33.3333333333333%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 50 0

[1] "Accuracy :66.6666666666667%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0 1 1.0000000

versicolor 0.5 1 0.6666667

virginica NaN 0 NaN

[1] "Kappa :0.5"

[1] 1

[1] "Number of instances :150"

[1] "Correctly Classified Instances :144 96%"

[1] "InCorrectly Classified Instances :6 4%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 49 1

virginica 0 5 45

[1] "Accuracy :96%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.0000000

versicolor 0.9074074 0.98 0.9423077

virginica 0.9782609 0.90 0.9375000

[1] "Kappa :0.94"

[1] 2

[1] "Number of instances :150"

[1] "Correctly Classified Instances :146 97.3333333333333%"

[1] "InCorrectly Classified Instances :4 2.66666666666667%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 47 3

virginica 0 1 49

[1] "Accuracy :97.3333333333333%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.0000000

versicolor 0.9791667 0.94 0.9591837

virginica 0.9423077 0.98 0.9607843

[1] "Kappa :0.96"

[1] 3

[1] "Number of instances :150"

[1] "Correctly Classified Instances :149 99.3333333333333%"

[1] "InCorrectly Classified Instances :1 0.666666666666667%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 1 49

[1] "Accuracy :99.3333333333333%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1.0000000 1.00 1.000000

versicolor 0.9803922 1.00 0.990099

virginica 1.0000000 0.98 0.989899

[1] "Kappa :0.99"

[1] 4

[1] "Number of instances :150"

[1] "Correctly Classified Instances :150 100%"

[1] "InCorrectly Classified Instances :0 0%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 0 50

[1] "Accuracy :100%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1 1 1

versicolor 1 1 1

virginica 1 1 1

[1] "Kappa :1"

[1] 5

[1] "Number of instances :150"

[1] "Correctly Classified Instances :150 100%"

[1] "InCorrectly Classified Instances :0 0%"

[1] "Comparision Matrix :"

predicted

setosa versicolor virginica

setosa 50 0 0

versicolor 0 50 0

virginica 0 0 50

[1] "Accuracy :100%"

[1] "Per-class Precision, Recall, and F-1 :"

precision recall f1

setosa 1 1 1

versicolor 1 1 1

virginica 1 1 1

[1] "Kappa :1"

[1] 6

Contribute

----------

- Source Code: <https://github.com/ippilisaikiran/EMSDM>

Support

-------

If you are having issues, please let us know.

We have a mailing located at: ippilisaikiran@gmail.com