Y.Harika
192325023

CSA1618 DWDM

EXPERIMENT-30

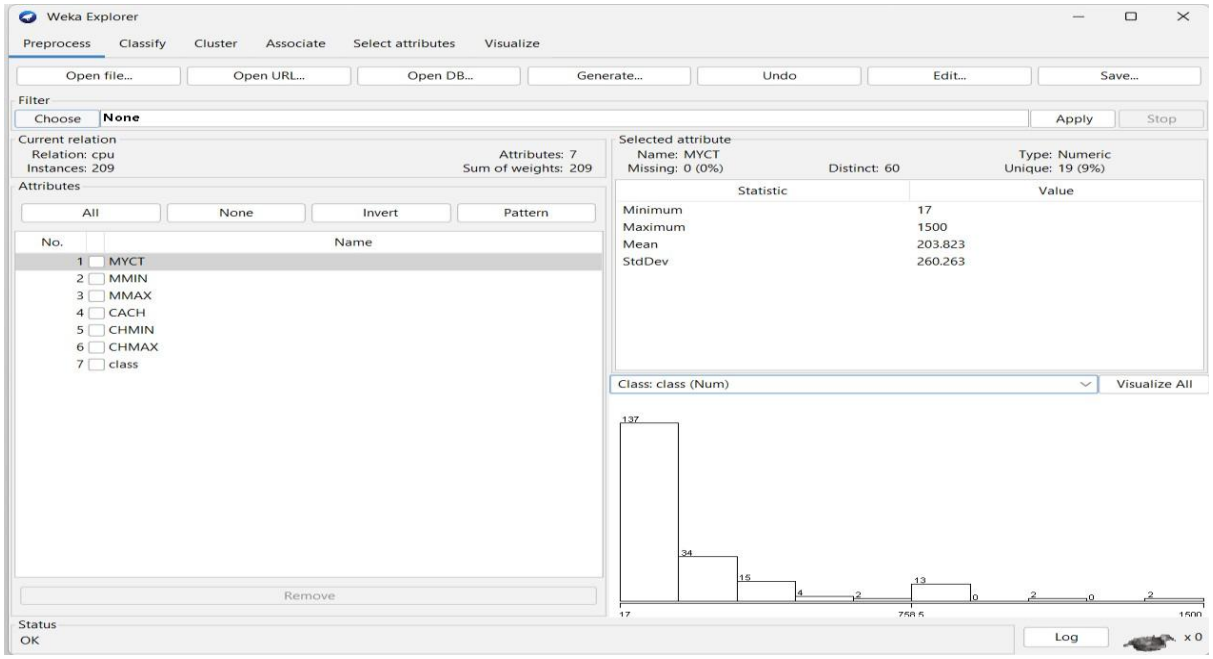
DESCRIPTION NUMERICAL PREDICTION ANALYSIS USING LINEAR REGRESSION THROUGH WEKA

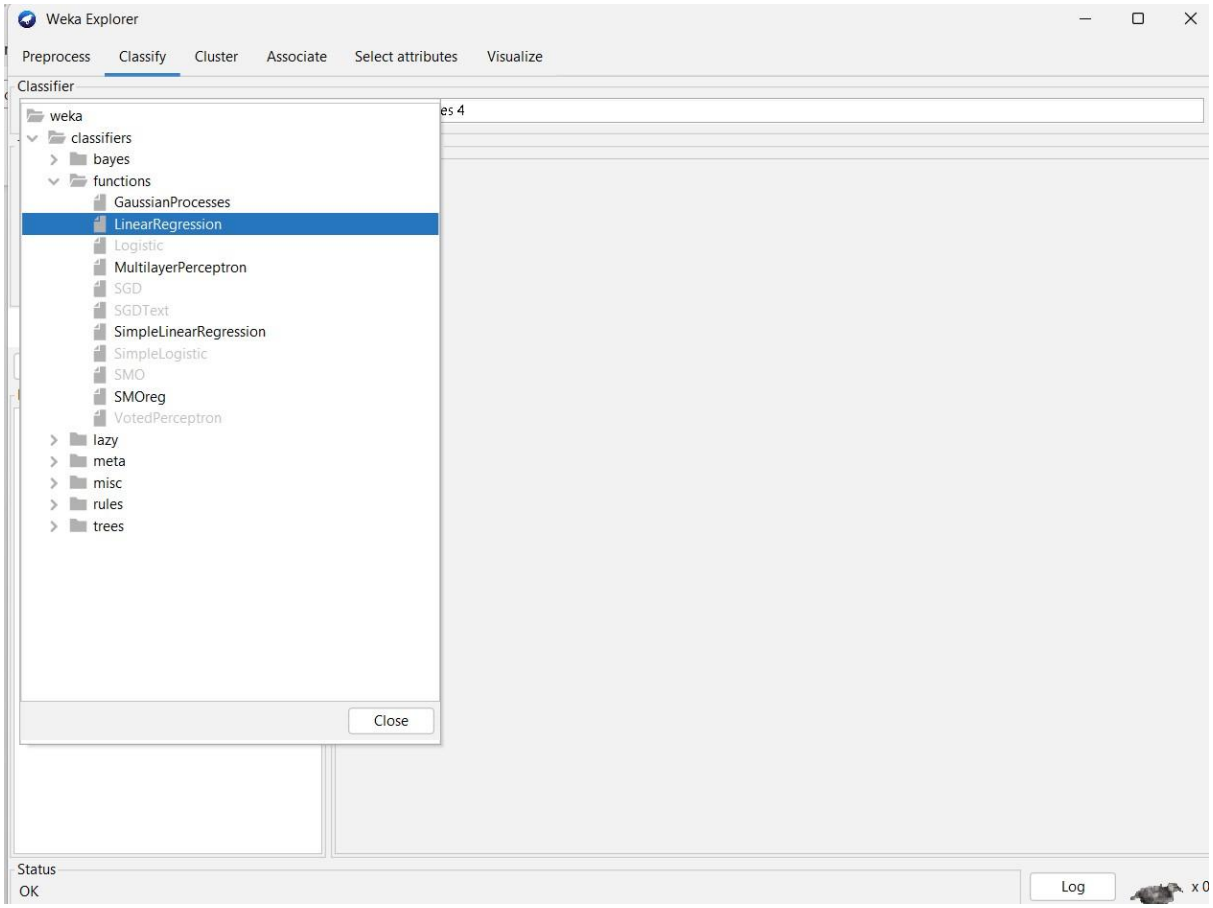
AIM:

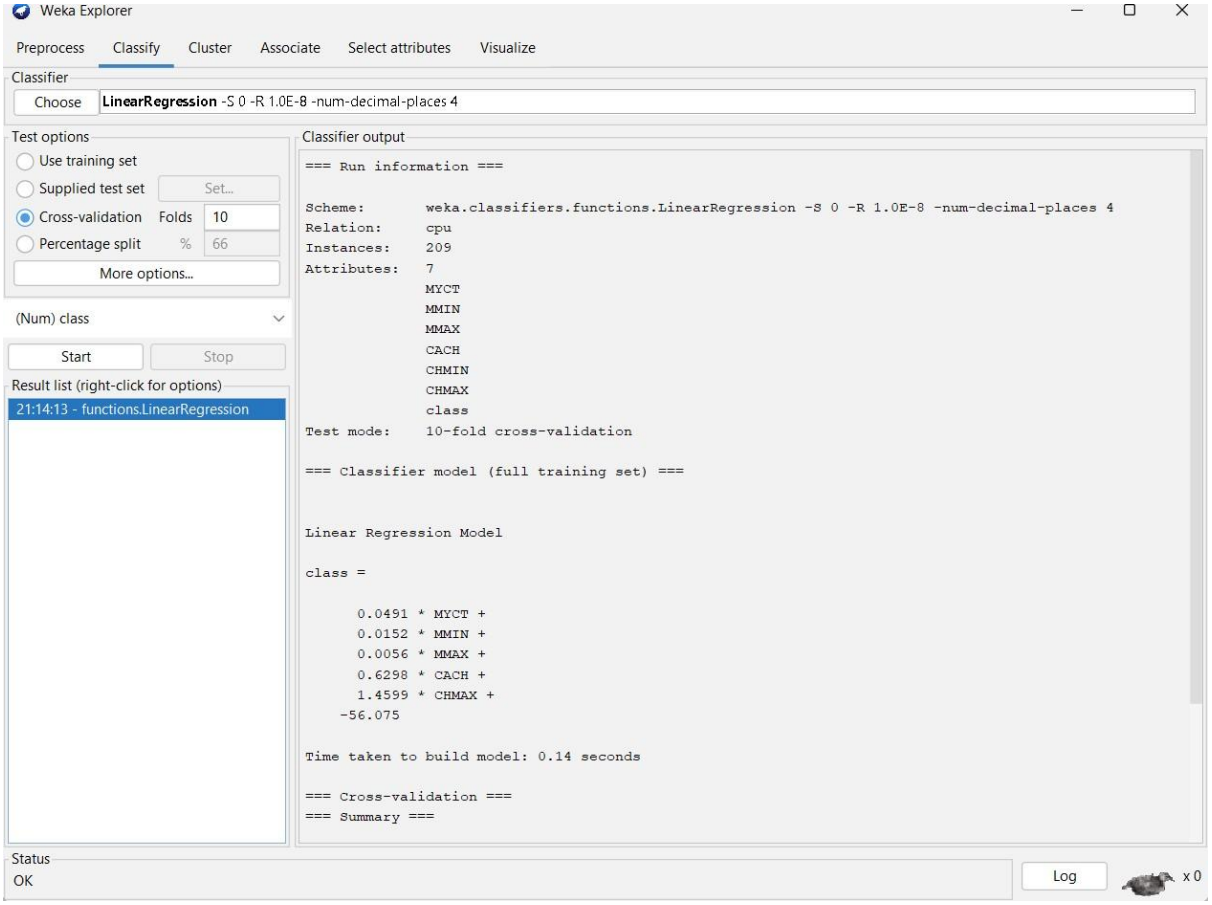
To create description numerical prediction analysis using linear regression through weka tool.

PROCEDURE:

1. Download and install WEKA.
2. Open WEKA and Choose "Explorer" from the main menu.
3. Under Preprocess, Click on the open file button and select the dataset. Ensure the target variable is numerical (continuous values like price, sales, etc.).
4. Go to the Classify tab.
5. Click and Choose Linear Regression.
6. Under Test options, choose: Cross-validation (10 Folds).
7. Click Start to begin classification.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: cpu
Instances: 209
Attributes: 7
Sum of weights: 209
Attributes
All
None
Invert
Pattern
No.
Name
1 MYCT
2 MMIN
3 MMAX
4 CACH
5 CHMIN
6 CHMAX
7 class
Remove
Selected attribute
Name: MYCT
Missing: 0 (0%)
Distinct: 60
Type: Numeric
Unique: 19 (9%)
Statistic
Value
Minimum 17
Maximum 1500
Mean 203.823
StdDev 260.263
Class: class (Num)
Visualize All
137
34
15
4
2
13
0
2
0
2
17
758.5
1500
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
GaussianProcesses
LinearRegression
Logistic
MultilayerPerceptron
SGD
SGDText
SimpleLinearRegression
SimpleLogistic
SMO
SMOreg
VotedPerceptron
lazy
meta
misc
rules
trees
Close
Status
OK
Log
x 0



OBSERVATION:

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Linear Regression Model class =

0.0491 * MYCT +

0.0152 * MMIN +

0.0056 * MMAX +

0.6298 * CACH +

1.4599 * CHMAX +

-56.075

Time taken to build model: 0.14 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9012
Mean absolute error	41.0886
Root mean squared error	69.556
Relative absolute error	42.6943 %
Root relative squared error	43.2421 %
Total Number of Instances	209

RESULT:

Thus, dataset has been observed using linear regression model. If the value of cross validation folds decreases time for creating model will be less than when folds value high, and the mean absolute error and Root mean square error values decreases with increase in the cross validation folds value.