

Ex. No: -3

Implement the machine
learning algorithm for the
dataset using weka tool

Nithishkumar, K
18C066

Aim

To validate the predication performance of multi dimensional data using machine learning model on weka tool

Tool description

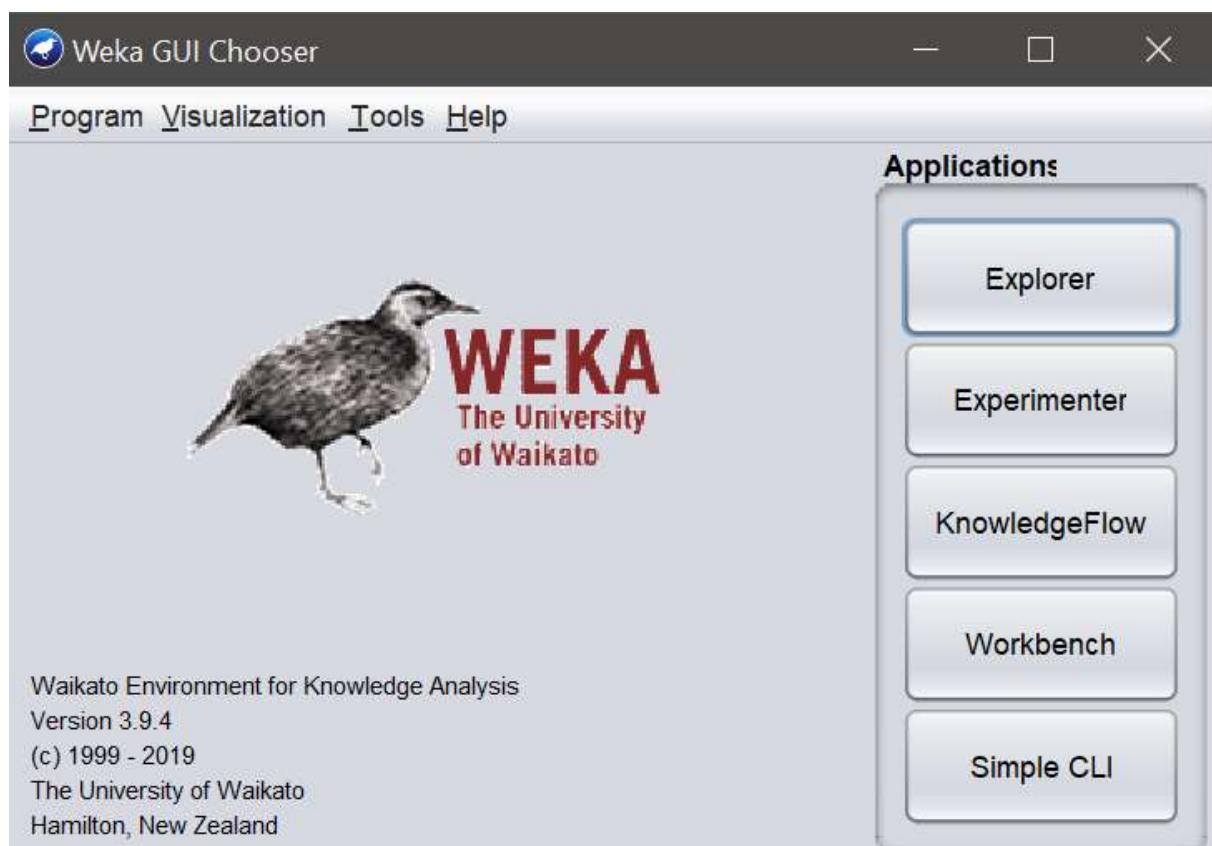
weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a data set or called from your own Java Code.

Execution steps

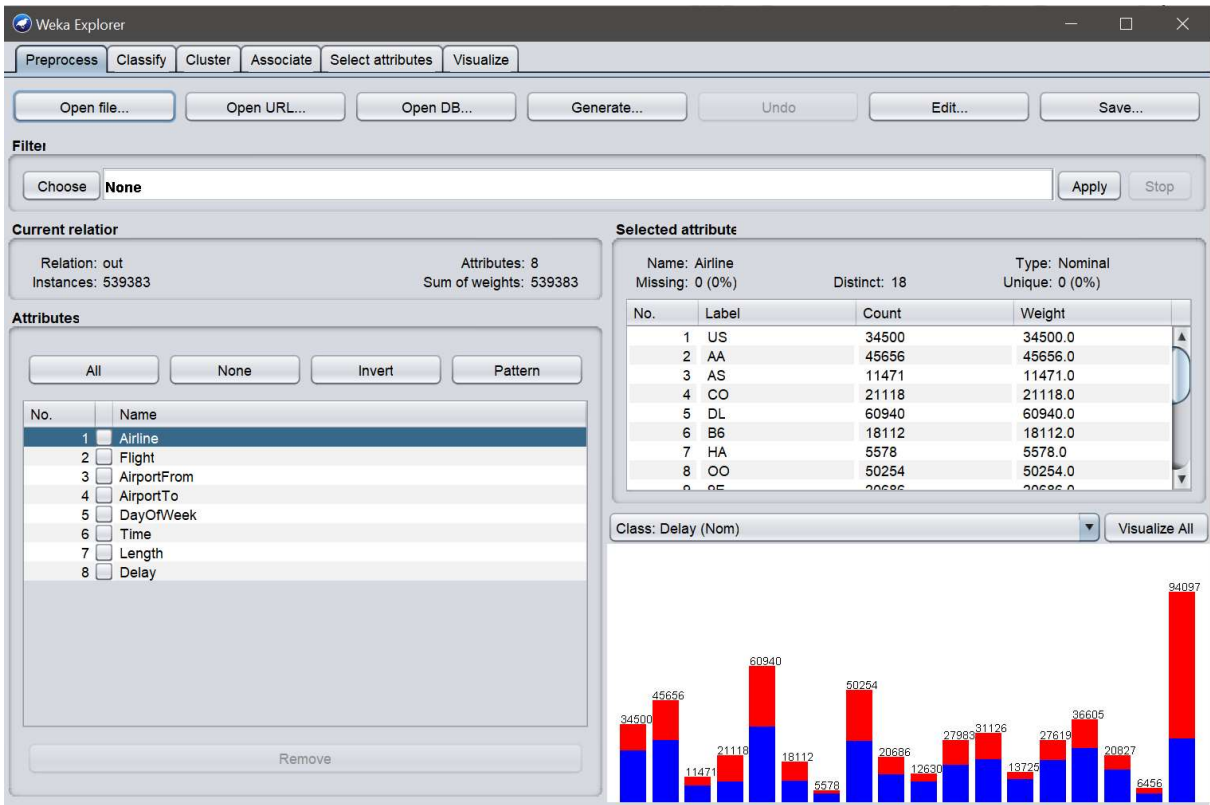
1. Download weka and Install. visit the weka download page and create a version of weka suitable for your Computer.
2. Start weka
3. open the data
4. select and Run an algorithm
5. Review Results.

OUTPUT:

1. Installation:



2. Loading Dataset:



3. Preprocessing Data:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **Normalize -S 1.0 -T 0.0** Apply Stop

Current relation
Relation: out
Instances: 539383
Attributes: 8
Sum of weights: 539383

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> Airline
2	<input type="checkbox"/> Flight
3	<input type="checkbox"/> AirportFrom
4	<input type="checkbox"/> AirportTo
5	<input type="checkbox"/> DayOfWeek
6	<input type="checkbox"/> Time
7	<input type="checkbox"/> Length
8	<input type="checkbox"/> Delay

Remove

Selected attribute
Name: Airline
Missing: 0 (0%)
Distinct: 18
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	US	34500	34500.0
2	AA	45656	45656.0
3	AS	11471	11471.0
4	CO	21118	21118.0
5	DL	60940	60940.0
6	B6	18112	18112.0
7	HA	5578	5578.0
8	OO	50254	50254.0
9	OE	20686	20686.0
10	UA	27968	27968.0
11	WN	31126	31126.0
12	SW	13728	13728.0
13	AL	27618	27618.0
14	EV	36605	36605.0
15	MO	20827	20827.0
16	44	6456	6456.0
17	94097	94097	94097.0

Class: Delay (Nom) Visualize All

Status

4. Apply Naïve Bayes Classifier:

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **NaiveBayes**

Test options

- ☐ Use training set
- ☐ Supplied test set Set...
- ☒ Cross-validation Folds **10**
- ☐ Percentage split % **66**

More options...

(Nom) Delay

Start Stop

Result list (right-click for options)

07:56:45 - bayes.NaiveBayes

Classifier output

```
==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances      341192           63.256 %
Incorrectly Classified Instances    198191           36.744 %
Kappa statistic                    0.2398
Mean absolute error                 0.426
Root mean squared error            0.4796
Relative absolute error             86.2201 %
Root relative squared error        96.5044 %
Total Number of Instances         539383

==== Detailed Accuracy By Class ====

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area
Weighted Avg.   0.633   0.398   0.630     0.633   0.625     0.245   0.668    0.663

==== Confusion Matrix ====

      a      b  <-- classified as
226392  72727 |      a = 0
125464 114800 |      b = 1
```

Status

5. Apply Simple KMeans Clustering:

Weka Explorer

Preprocess | **Cluster** | Associate | Select attributes | Visualize

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -l1 -1.25 -l2 -1.0 -N 2 -A "weka.core.EuclideanDistance" -R first-last" -I 500 -num-slot:

Cluster mode

- ☒ Use training set
- ☐ Supplied test set Set...
- ☐ Percentage split % **66**
- ☐ Classes to clusters evaluation

(Nom) Delay

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

07:59:19 - SimpleKMeans

Clusterer output

```
Final cluster centroids:

Attribute      Full Data      Cluster#
(539383.0)    (247360.0)    (292023.0)

=====
Airline        WN            WN            OO
Flight         0.3106       0.2237       0.3842
AirportFrom    ATL          ATL          ATL
AirportTo      ATL          ATL          ATL
DayOfWeek      4            6            3
Time           0.5547       0.579        0.5342
Length         0.2018       0.2136       0.1919
Delay          0            1            0

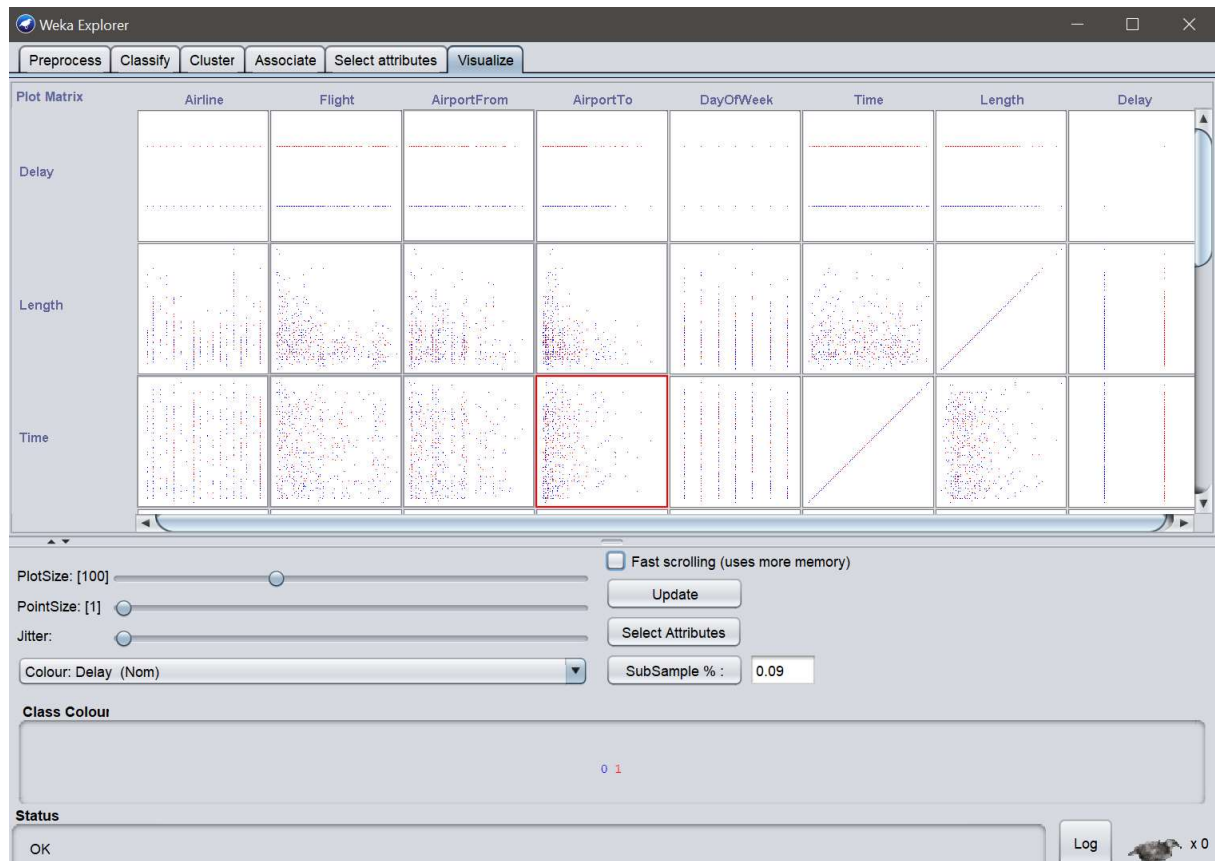
Time taken to build model (full training data) : 1.5 seconds

==== Model and evaluation on training set ====

Clustered Instances

0      247360 ( 46%)
1      292023 ( 54%)
```

6. Scatter plot Visualization:



Result:

Thus , the machine learning algorithm for the dataset using weka tool was implemented.