

Experiment1.1

Student Name: Himanshu

UID: 20BCS7944

Branch: CSE

Section: 905/A

Semester: 6

Date of Performance: 15/02/2023

Subject Name: Data Mining Lab

Subject Code: 20CSP-376

1) **Aim:**

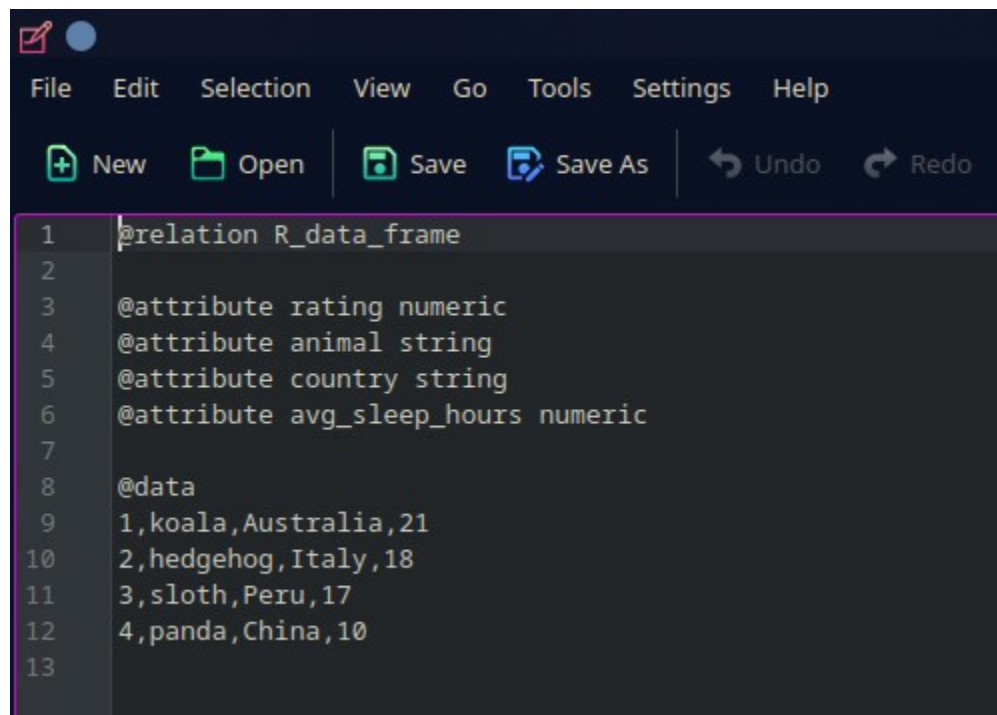
Demonstration of preprocessing on .arff file using R Programming.

2) **Objective:**

To represent the creation of file using R Studio and displaying the pattern on Weka Tool for further extraction and analysis of knowledge.

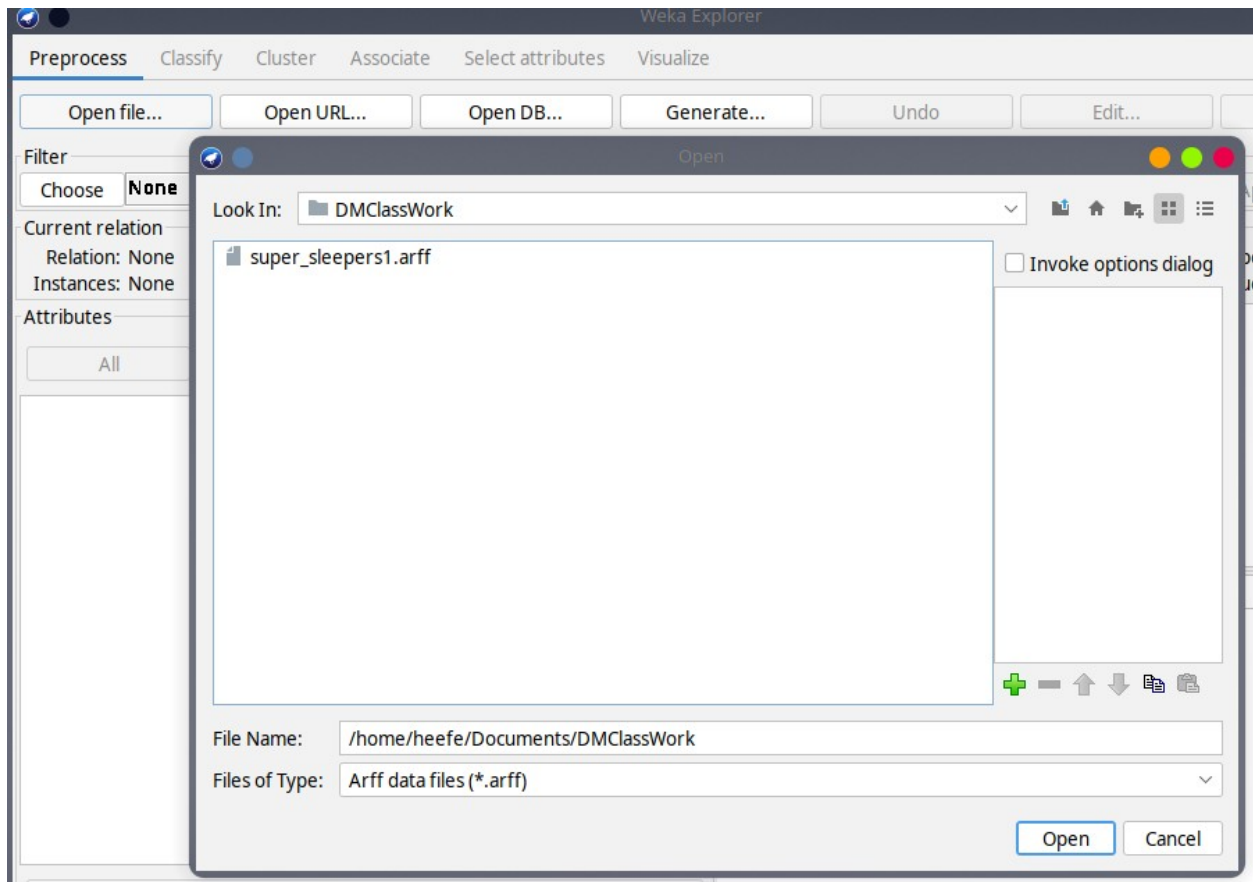
3) **Code and Output:**

- Created a .arff file as -



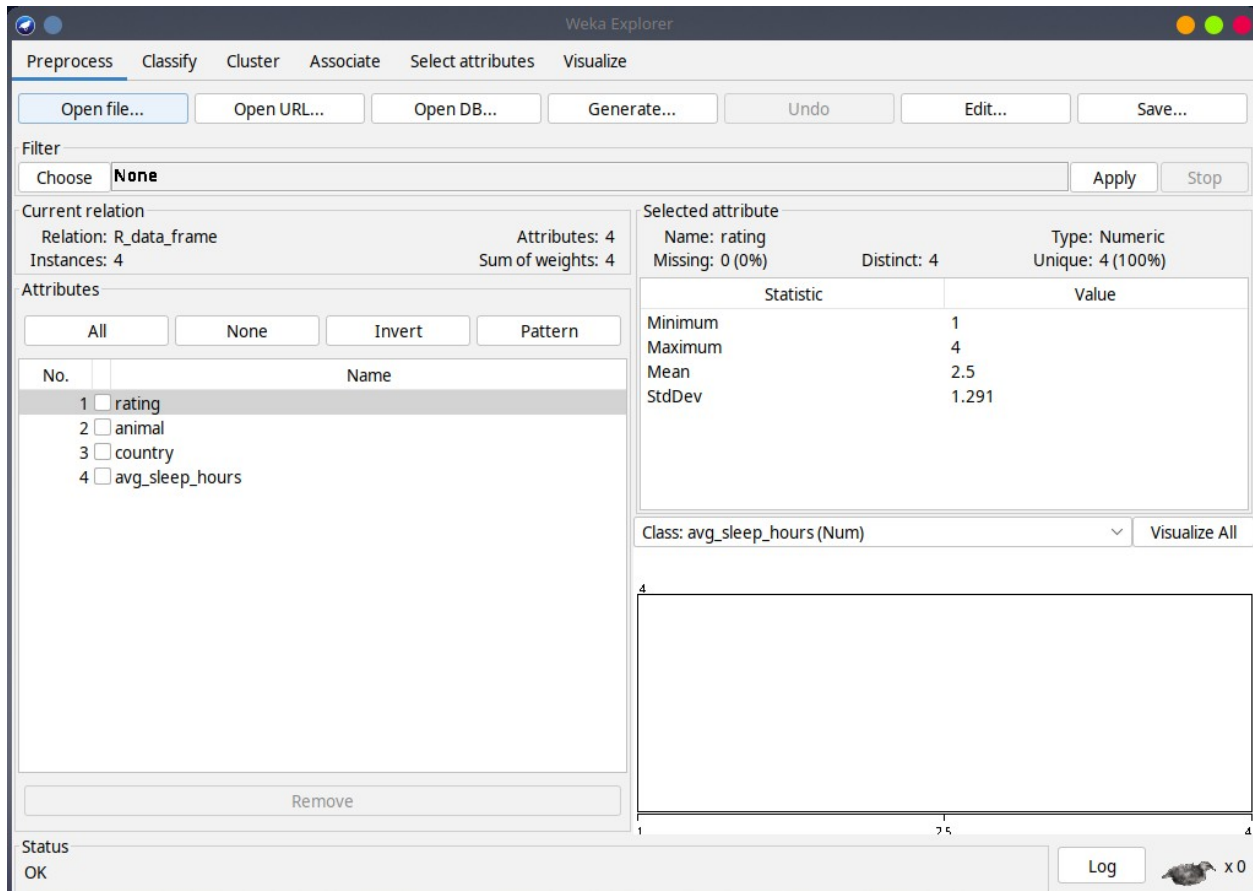
```
1 @relation R_data_frame
2
3 @attribute rating numeric
4 @attribute animal string
5 @attribute country string
6 @attribute avg_sleep_hours numeric
7
8 @data
9 1,koala,Australia,21
10 2,hedgehog,Italy,18
11 3,sloth,Peru,17
12 4,panda,China,10
13
```

- Opened the file using Rweka -



```
5 source( "/home/heefe/Documents/DMClassWork/" ,
6
7 #checking working directory
8 getwd()
9
0 #Creating Data frame
1 rating <- 1:4
2 animal <- c('koala','hedgehog','sloth','panda')
3 country <- c('Australia','Italy','Peru','China')
4 avq sleep hours <- c(21,18,17,10)
```

Extracted information -



The screenshot displays the Weka Explorer application window. The 'Preprocess' tab is active, showing options to open files, URLs, or databases, and buttons for generating, undoing, editing, and saving. The 'Filter' section is set to 'None'. The 'Current relation' is 'R_data_frame' with 4 attributes and 4 instances. The 'Attributes' list includes 'rating', 'animal', 'country', and 'avg_sleep_hours'. The 'Selected attribute' section shows statistics for 'rating': Minimum (1), Maximum (4), Mean (2.5), and StdDev (1.291). The 'Class' is set to 'avg_sleep_hours (Num)'. A 'Visualize All' button is present. The status bar at the bottom shows 'OK' and a 'Log' button.

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: R_data_frame Attributes: 4
Instances: 4 Sum of weights: 4

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> rating
2	<input type="checkbox"/> animal
3	<input type="checkbox"/> country
4	<input type="checkbox"/> avg_sleep_hours

Remove

Selected attribute
Name: rating Type: Numeric
Missing: 0 (0%) Distinct: 4 Unique: 4 (100%)

Statistic	Value
Minimum	1
Maximum	4
Mean	2.5
StdDev	1.291

Class: avg_sleep_hours (Num) Visualize All

Status
OK Log x 0