



Experiment1.1

Name: Rahul Maurya

Branch: B.E-CSE

Semester: 6th

Subject Name: Data Mining Lab

UID: 20BCS7260

Section/Group: 716/A

Date of Performance: 15/02/2023

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:

Program 1

#Experiment 1

#arff - attribute relation file format

```
#getting and setting the current directory
```

```
getwd()
```

```
setwd("write your directory")
```

```
getwd()
```

```
library("RWeka")
```

```
#Creation of DataFrame
```

```
UID<- 1:5
```

```
NAME<-c('Rahul','Sham','Ram','Rohit','Geeta')
```

```
AGE<-c(21,21,20,19,21)
```

```
GENDER<-c('M','M','M','M','F')
```

```
SECTION<-c(54,54,54,54,54)
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
COURSE<-c('CSE','CSE','CSE','CSE','CSE')
```

```
SUBJECT1<-c(95,94,84,89,90)
```

```
SUBJECT2<-c(94,94,84,89,92)
```

```
SUBJECT3<-c(93,94,82,92,93)
```

```
TOTAL<-SUBJECT1+SUBJECT2+SUBJECT3
```

```
stud_data <-
```

```
data.frame(UID,NAME,AGE,GENDER,SECTION,COURSE,SUBJECT1,SUBJECT2,SUBJECT3,TOTAL,stringsAsFactors = FALSE)
```

```
stud_data
```

```
#make sure to set dstring as factors to false.
```

```
#so that string values are stored as characters and not vectors.
```

```
#so that our data will not be considered as categorical
```

```
str(stud_data)
```

```
#creating arff file
```

```
write.arff(stud_data,file("stud_data.arff"))
```

```
N=read.arff("stud_data.arff")
```

```
print(N)
```

```
N
```

```
head(N,2)
```

OUTPUT

	UID	NAME	AGE	GENDER	SECTION	COURSE	SUBJECT1	SUBJECT2	SUBJECT3	TOTAL
1	1	Rahul	21	M	54	CSE	95	94	93	282
2	2	Sham	21	M	54	CSE	94	94	94	282
3	3	Ram	20	M	54	CSE	84	84	82	250
4	4	Rohit	19	M	54	CSE	89	89	92	270
5	5	Geeta	21	F	54	CSE	90	92	93	275
> N										
	UID	NAME	AGE	GENDER	SECTION	COURSE	SUBJECT1	SUBJECT2	SUBJECT3	TOTAL
1	1	Rahul	21	M	54	CSE	95	94	93	282
2	2	Sham	21	M	54	CSE	94	94	94	282



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

📁 > This PC > DATA (D:) > sem 6 > AD

Name	Date modified	Type	Size
AndroidStudioEssentialsPreview	13-02-2023 10:44	Microsoft Edge PD...	2,740 KB
AndTutorials-3_1-CC	13-02-2023 10:44	Microsoft Edge PD...	6,298 KB
ex1	22-02-2023 10:44	R Source File	1 KB
stud_data.arff	22-02-2023 10:37	ARFF File	1 KB

stud_data.arff - Notepad

File Edit View

```
@relation R_data_frame

@attribute UID numeric
@attribute NAME string
@attribute AGE numeric
@attribute GENDER string
@attribute SECTION numeric
@attribute COURSE string
@attribute SUBJECT1 numeric
@attribute SUBJECT2 numeric
@attribute SUBJECT3 numeric
@attribute TOTAL numeric

@data
1,Rahul,21,M,54,CSE,95,94,93,282
2,Sham,21,M,54,CSE,94,94,94,282
3,Ram,20,M,54,CSE,84,84,82,250
4,Rohit,19,M,54,CSE,89,89,92,270
5,Geeta,21,F,54,CSE,90,92,93,275
```