



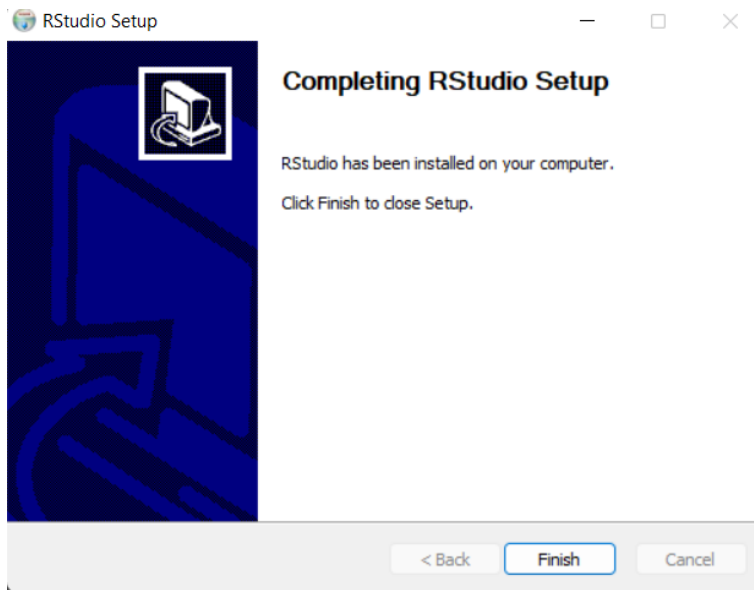
EXPERIMENT -1

NAME- Jatin	UID- 20BCS5951
BRANCH- BE- CSE	SECTION- DM 605 B
SEMESTER- 6 th	DATE- 18/02/2023
SUBJECT- Data Mining Lab	SUBJECT CODE- 20CSP-376

- **Aim .** Demonstration of preprocessing on .arff file using student data .arff.
- **Objective:**
Learning about
ARFF files and
how to create
ARFF File
(Attribute
relation File
Format)
- **Installation:**
 - To install R, go to cran.r-project.org
 - Depending on your operating system, click Download R for (your operating system).
 - Click on install R for the first time.
 - Click Download R for Windows. Open the downloaded file.
 - Select the language you would like to use during the installation. Then click OK.
 - Click Next.
 - Select where you would like R to be installed. It will default to your Program Files on your C Drive. Click Next.
 - You can then choose which installation you would like.
 - (Optional) If your computer is a 64-bit, you can choose the 64-bit User Installation. Then click Next.
 - Then specify if you want to customized your startup or just use the defaults. Then click Next.
 - Then you can choose the folder that you want R to be saved within or the default if the R folder that was created. Once you have finished, click Next.

- You can then select additional shortcuts if you would like. Click Next.
- Click Finish.
- Next, download RStudio. Go to www.rstudio.com
- Click Download RStudio.
- The RStudio installation wizard will pop-up. Click Next and go through the installation steps.
- Congratulations! You have now installed R and RStudio.

- **Installation Screenshots**



R Choose R Installation X

RStudio requires an existing installation of R.

Please select the version of R to use.

- ☒ Use your machine's default 64-bit version of R
☐ Use your machine's default 32-bit version of R
☐ Choose a specific version of R:

[64-bit] C:\Program Files\R\R-4.2.2

You can also customize the rendering engine used by RStudio.

Rendering Engine: Auto-detect (recommended) ▼

Browse...

OK

Cancel

- **Script and Output:**

```
exp (-200)
log (100,base=10)
Actual<-runif (100, min=1, max=5)
Actual
head (Actual)
head (Actual, 10)
head (Actual, 20)
Predicted<-runif (100,min=1, max=5)
Predicted
head (predicted)
Actual-Predicted
```



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
sampledata.csv x
1 "Actual", "Predicted"
2 4.29059976246208, 4.55274042859674
3 3.82663166802377, 3.21122537087649
4 1.97292643040419, 1.88762484770268
5 4.09614888671786, 3.82395818363875
6 4.5990875903517, 4.47623041737825
7 1.72047721594572, 1.84274540655315
8 3.78937337826937, 4.35565521381795
9 2.56729054544121, 1.7249355353415
10 4.55651433579624, 2.28252208512276
11 4.0620190910995, 3.51531759370118
12 1.16050578746945, 1.66128084994853
13 4.43802418187261, 4.60965051222593
14 3.21274121291935, 1.40312664117664
15 1.10545007605106, 3.10497531760484
16 2.93034338951111, 1.77734740078449
17 2.00400864426047, 1.41477120667696
18 3.29777921270579, 1.42903277277946
19 1.36517356149852, 3.5874051572755
20 3.03937948308885, 4.77216534130275
21 4.54521915595978, 3.66475276276469
22 4.49685949552804, 2.43497786298394
23 2.24077562335879, 2.6171507550005
1:1 Text file
Console Terminal Background Jobs
R 4.2.2 ~ /
[41] 4.648586 4.211088 3.524659 4.858330 1.891832 2.478825 1.474051 3.950407 1.059630 4.684877
[51] 2.599054 2.602627 2.203042 2.988405 1.383619 3.194187 2.456303 1.667271 2.321240 2.707614
[61] 4.873318 1.286881 4.874897 2.479204 2.246342 3.720380 4.667059 2.168207 2.352631 2.762163
[71] 4.656750 1.736153 4.995815 2.331094 4.635334 2.649968 2.424574 1.933011 4.638529 4.652509
[81] 3.430559 4.959543 1.956682 2.928356 4.418352 1.069551 2.367337 1.894997 4.671658 3.291052
[91] 1.694885 2.605009 2.352893 1.534394 2.860119 2.870131 2.613283 1.506910 4.782045 1.280210
> head(Predicted)
[1] 4.552740 3.211225 1.887625 3.823958 4.476230 1.842745
> Actual-Predicted
[1] -0.26214067 0.61540630 0.08530158 0.27219070 0.12285717 -0.12226819 -0.56628184
[8] 0.84235501 2.27399225 0.54670150 -0.50077506 -0.17162633 1.80961457 -1.99952524
[15] 1.15299599 0.58923744 1.86874644 -2.22223160 -1.73278586 0.88046639 2.06188163
[22] -0.37638413 0.53240035 1.89404950 -0.93165940 2.36763063 0.52111662 -0.63889100
[29] -1.27166768 0.27452493 0.55279501 -0.58380958 -0.54780159 -2.51576753 -0.13478527
[36] 1.60595711 1.94135410 0.05757740 -2.35103873 -1.70175367 -3.59370166 -1.99237521
[43] -1.86431880 -0.32119039 3.02499106 -0.31702061 1.48708515 -0.52132253 3.07338615
```

- **Preprocessing
on .arff file
using student
data .arff**

```
rating<-5:1
name<-c('Jatin','Himanshu','Rahul','Aryan','Tyagi')
uid<-c('20BCS5951','20BCS9655','20BCS9883','20BCS9825','20BCS9531')
marks<-c(78,65,75,98,45)
student_info<-data.frame(rating,name,uid,marks,stringasfactors=TRUE)
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
sampledata.csv x studentdata.csv x studentdata1.csv x
1 "", "rating", "name", "uid", "marks", "rating.1", "name.1", "uid.1", "marks.1", "stringasfactors"
2 "1", 5, "Jatin", "20BCS5951", 78, 5, "Jatin", "20BCS5951", 78, TRUE
3 "2", 4, "Himanshu", "20BCS9655", 65, 4, "Himanshu", "20BCS9655", 65, TRUE
4 "3", 3, "Rahul", "20BCS9883", 75, 3, "Rahul", "20BCS9883", 75, TRUE
5 "4", 2, "Aryan", "20BCS9825", 98, 2, "Aryan", "20BCS9825", 98, TRUE
6 "5", 1, "Tyagi", "20BCS9531", 45, 1, "Tyagi", "20BCS9531", 45, TRUE
7
```



Project: (None) ▾

Environment

History

Connections

Tutorial

Import Dataset

169 MiB

R

Global Environment

student_info

5 obs. of 5 variables

Values

Actual

num

[1:100]

4.29 3.83 1.97 4.1 4.6 ...

marks

num

[1:5]

78 65 75 98 45

name

chr

[1:5]

"Jatin" "Himanshu" "Rahul" "Aryan" "Tyagi"

Predicted

num

[1:100]

4.55 3.21 1.89 3.82 4.48 ...

rating

int

[1:5]

5 4 3 2 1

uid

chr

[1:5]

"20BCS5951" "20BCS9655" "20BCS9883" "20BCS9825"...