

Data Analytics in Abroad and Indian Education System-Using Data Mining Classification Techniques by R Language

J Uma Mahesh, N Chandrakanth, M Ravinder Reddy

Department of CSE, Geethanjali College of Engineering & Technology, Hyderabad, India

E-mail: umamaheshsjcet@gmail.com, chandra.nagulapalli@gmail.com, ravinder0512@gmail.com

Abstract

Education System in recent years has been a progression, in Indian and Abroad Education system. In selecting the next education establishment by the scholars. The most key terms of selecting associate in nursing institute area unit pursue data, institute enfranchisement, institute ranking, freshman retention, graduation rates and strength of the college resources, location, feel of field life, placement records, analysis activities, course length, course outcome, educational offerings, activities and sports, price of the provision of economic aid and etc. This paper proposes to handle the coed quality in choosing an establishment to pursue educational activity in abroad/India supported the on top of mentioned key terms by having a deep analysis visualization data processing, classification and prediction model techniques visualization R language with Rattle Package.

Keywords: *Data mining, R-language, rattle package, classification-techniques, decision tree*

INTRODUCTION TO R AND

RATTLE PACKAGE

R Language

R is a powerful language and surroundings for sta- tistical computing and graphics. It is a public do- main (a thus referred to as \GNU") project that is analogous to the industrial S language and surroundings that was developed at Bell Laboratories (for-merely AT&T, currently aglow Technologies) by John Chambers and

colleagues are often thought of as a unique implementation of S, and is way utilized in as an academic language and analysis tool. The main advantages of R are the fact that R is freeware and that there is a lot of help available online [1, 2]. It is quite just like alternative programming packages akin to Mat workplace (not freeware), however, a lot of easy than programming languages akin to C++ or FORTRAN you will use R because it is, except for instructional

functions we have a tendency to value more highly to use R in combination with the Studio interface (also freeware), that has Associate in Nursing organized layout and a number of other further choices. R uses data processing Techniques and classification techniques [3, 4].

Rattle Package

The aim is to supply an easy and intuitive interface that enables a user to quickly load knowledge from a CSV file (or via ODBC), remodel and explore the information, build and appraise models, and export models as PMML (predictive modeling markup language) or as scores. All of this with knowing little about R. All R commands are logged and commented through the log tab. Thus they are available to the user as a script file or as an aide for the user to learn R or to copy-and-paste directly into R itself. Rattle also exports a number of utility functions and the graphical user interface, invoked as `rattle ()`.

Data Mining with R

This series aims to capture new developments and applications in data processing and information discovery, whereas, summarizing the process tools and techniques helpful in knowledge analysis. This series encourages the

mixing of mathematical, applied mathematics, and process strategies and techniques through the publication of a broad vary of textbooks, reference works, and handbooks. The inclusion of concrete examples and applications is very inspired. The scope of the series includes, however, is not restricted to, titles within the areas of knowledge of mining and knowledge discovery ways and applications, modeling, algorithms, theory and foundations, information and information image, data processing systems and tools, and privacy and security problems that permits the classification techniques [5, 6].

Data Mining with R we can do many things:

- Data Exploration/Knowledge Exploration
- Decision Trees/Classification.
- k-means Clustering and Hierarchical Clustering.
- Outlier Detection.
- Time Series Decomposition and Forecasting.
- Time Series Clustering and Classification.
- Association Rules.
- Text Mining
- Social Network Analysis
- Multidimensional Scaling (MDS) and etc.

Related Work using R

Case Study: Predicting Algae Blooms (Refer Data Mining with R)

This case study will introduce you to some basic tasks of data mining: data Pre-processing, exploratory data analysis and predictive model construction [7, 8]. For this initial case study we have selected a small problem by data mining Standards. Namely, we are addressing the problem of predicting the frequency Occurrence of several harmful algae in water samples. By this related work we got an idea to analyze Abroad Education and Indian Education System-Using Data Mining Classification and Prediction Model Techniques by R Language

R-Console

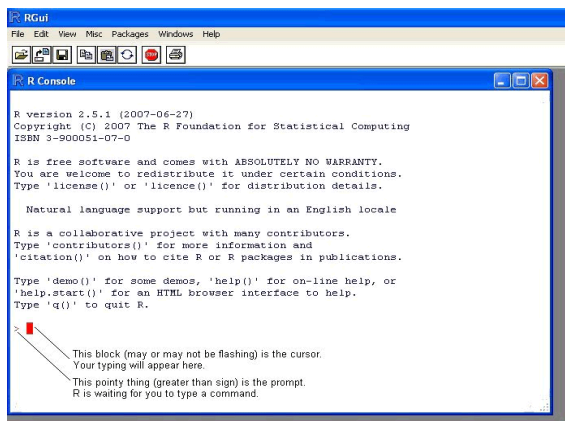


Fig. 1: R-Console.

R Console Examples

R Session

After R is started, there is a console waiting for input. At the prompt (>), you

can enter numbers and perform calculations.

```
> 1 + 2 [1] 3
```

Variable Assignment

We assign values to variables with the assignment operator "=". Just typing the variable by itself at the prompt will print out the value. We should note that another form of assignment operator "<-" is also in use.

```
> x = 1 > x [1] 1
```

Functions

R functions are invoked by its name, then followed by the parenthesis, and zero or more arguments. The following apply the function c to combine three numeric values into a vector.

```
> c(1, 2, 3) [1] 1 2 3
```

Comments

All text after the pound sign "#" within the same line is considered a comment.

```
> 1 + 1 # this is a comment
```

```
[1] 2
```

Extension Package

Sometimes we want extra practicality on the far side those offered by the core R library. So, as to put in associate degree extension package, you

must invoke the install packages operate at the prompt and follow the instruction.

```
> install.packages ()
```

Getting Help

R provides extensive documentation. For example, entering `? c` or `help(c)` at the prompt gives documentation of the function `c` in R. Please give it a try.

```
> help(c)
```

If you are not sure about the name of the function you are looking for, you can perform a fuzzy search with the `apropos` function.

```
> apropos                                ("nova")
[1] "anova"                                "anova.glm"
```

METHODOLOGIES

Classification Techniques

Decision Tree

The decision tree could be a structure that features root node, branch and leaf node. Every internal node denotes a check on attribute; every branch denotes the result of check and every leaf node holds the category label. The top node within the tree is that the root node.

R with Rattle Package

The aim is to supply a straightforward and intuitive interface that enables a user to quickly load knowledge from a CSV file

(or via ODBC), rework and explore the information, build and appraise models, and export models as PMML (predictive modeling markup language) or as scores. All of this with knowing very little concerning R. All R commands are logged and commented through the log tab. Thus they are available to the user as a script file or as an aide for the user to learn R or to copy-and-paste directly into R itself. Rattle also exports a number of utility functions and the graphical user interface, invoked as `rattle ()`.

R-Statistics

R and its libraries implement a good style of applied mathematics and graphical techniques, together with linear and nonlinear modeling, classical applied mathematics tests, time-series analysis, classification, clustering, and others. R is definitely protractible through functions and extensions, and, therefore, the R community is noted for its active contributions in terms of packages.

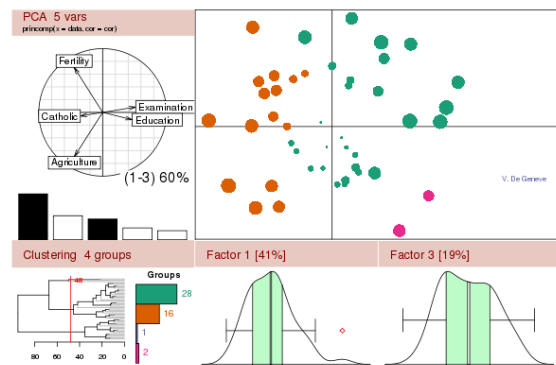


Fig. 2: R-Statistics.

Statistics/Results on Abroad and Indian Education

```
> university<-1:15
> institutename<-c("Alabama A&M
university","California institute of
Technology","university of
Datgton","fairfield university","East
carolina university")
> instituteaccreditation<-
c("SACS","WASC","AACSB","NACTE",
"NACTE")
> instituterating<-c(4.6,4.6,4.5,4.5,4.9)
> freshmanretention<-
c("average","average","good","average","g
ood")
> graduationrates<-
c("good","average","good","moderate","be
st")
> strengthoffacultyresources<-
c("good","good","yes","good","yes")
> Location<-
c("Alabama","California","Ohio","Connect
icut","North Carolina")
> FeelOfCampusLife<-
c("good","good","good","moderate","good
")
> PlacementRecords<-
c("moderate","yes","best","yes","best")
> ResearchActivities<-
c("yes","yes","yes","yes","yes")
> Courseduration<-c(2,2,2,2,2)
```

```
>courseoutcome<-
c("success","success","success","success",
"success")
>AcademicOfferings<-
c("M.S,B.SC","M.S,B.SC","M.S,B.SC","
M.S","M.S,B.SC")
>ActivitiesandSports<-
c("yes","yes","yes","yes","yes")
> fee<-
c(17968,12431,14184,27000,17951)
>df1<-
data.frame(university,institutename,institut
eaccreditation,instituteranking,freshmanrete
nsion,graduationrates,strengthoffacultyres
ources,Location,FeelOfCampusLife,Place
mentRecords,ResearchActivities,Coursedu
ration,courseoutcome,AcademicOfferings,
ActivitiesandSports,fee)
> names(df1)<-c("university
int","institutename
char","instituteaccreditation
char","instituterating
int","freshmanretention
char","graduationrates
char","strengthoffacultyresources
char","Location char","FeelOfCampusLife
char","PlacementRecords
char","ResearchActivities
char","Courseduration
int","courseoutcome
char","AcademicOfferings
char","ActivitiesandSports char","fee int")
```

```
>
write.csv(df1,"c:/university2/university4.csv")
>
df2<-
read.csv("c:/university2/university4.csv")
> print(df2)
//which prints data in .csv
>attach(test.data)

>pie(table(test.data$instituterating))
```

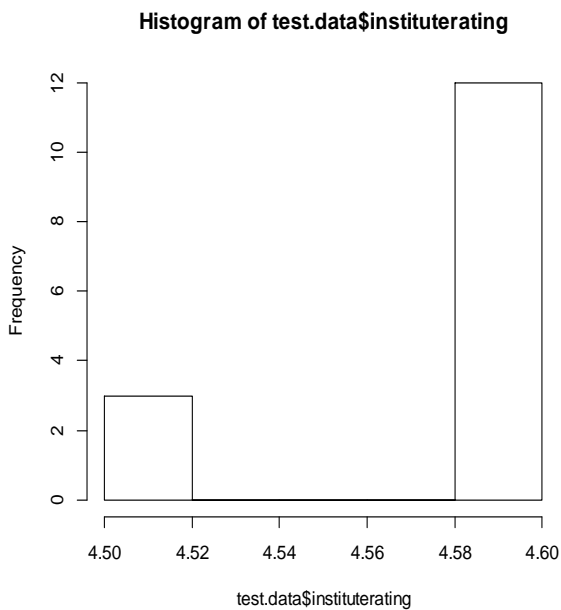
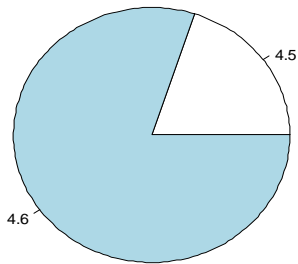


Fig. 3: Histogram.

```
>hist(test.data$instituterating)
```

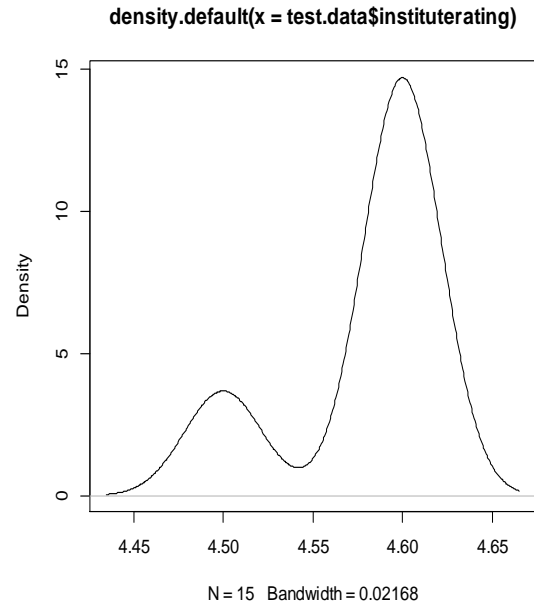


Fig. 4: Density.

```
>plot(density(test.data$instituterating))
```

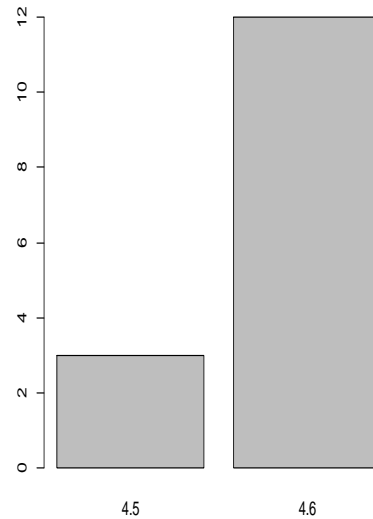


Fig. 5: Barplot.

```
>barplot(table(test.data$instituterating))
```

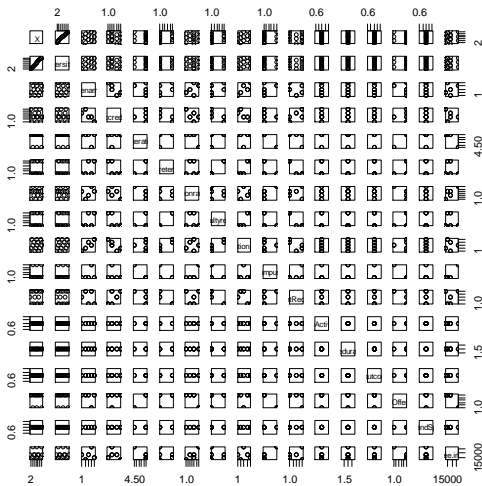
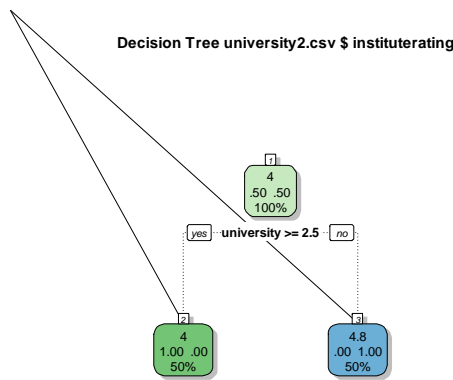


Fig. 6: Pairs.

>Pairs(test.data)

Decision Tree



Rattle 2015-Feb-28 11:22:28 UmaMahesh

Fig. 7: Decision Tree.

CONCLUSION

By applying a Classification algorithm like J48 and other R statistics we can predict and we can have a deep analysis over the abroad and Indian education system. The above results which has been generated by using R statistics and Rattle package.

Which shows student mobility in education system while selecting a top institution with all facilities? By this students will get an idea about which institution is best to continue his higher education. It is done by some form of statistical analysis and computing. In the next paper we are going to compare how best classifier algorithm can analyses the education system for selecting a top institution in abroad and Indian education, i.e., using J48, ID3, CART, BAYESIAN, etc.

REFERENCES

1. Adler, J. R in a Nutshell. O’Reilly, Sebastopol, CA; 2009.
2. Aknouche, A., Bibi, A. Quasi-maximum likelihood estimation of periodic garch and periodic arma-garch processes. *Journal of Time Series Analysis*. 2009; 30(1): 19–46p.
3. Beran, J., Whitcher, B., Maechler, M. Longmemo: Statistics for Long-Memory Processes. URL <http://CRAN.R-project.org/package=longmemo>. 2009.
4. Cook, D., Swayne, D. F. Interactive dynamic graphics for data analysis R. *Springer-Verlag*, New York; 2007.
5. Available at: www.rdatamining.com.
6. Available at: www.r-project.org/.

7. Available at:
en.wikipedia.org/wiki/Mobile_app.
8. Data Mining: Concepts and Techniques. Second Edition. Jiawei Han and Micheline Kamber.