VISVESVARAYA TECHNOLOGICAL UNIVERSITY,
BELAGAVI 590018

Report on

"PREDICTION OF DEMENTIA USING MRI DATASET"

By

AISHWARYA B (1BM14CS012)
ANITHA D M (1BM14CS018)
ASHIKA M (1BM14CS023)

Under the Guidance of

NAGARATHNA N
Associate Professor,
Department of CSE
BMS College of Engineering

Work carried out at

Department of Computer Science and Engineering
BMS College of Engineering
(Autonomous college under VTU)
P.O. Box No.: 1908, Bull Temple Road, Bangalore-560 019
2017-2018

BMS COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the DATA SCIENCE USING R Assignment titled "**PREDICTION OF DEMENTIA USING MRI DATASET"** has been carried out by AISHWARYA B (1BM14CS012), ANITHA D M (1BM14CS018) and ASHIKA M (1BM14CS023) during the academic year 2017-2018.

Signature of the guide
NAGARATHNA N
Associate Professor,
Department of Computer Science and Engineering
BMS College of Engineering, Bangalore

Signature of the Examiner

1.

2.

BMS COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We, AISHWARYA B (1BM14CS012), ANITHA D M (1BM14CS018) and ASHIKA M (1BM14CS023), students of 7th Semester, B.E, Department of Computer Science and Engineering, BMS College of Engineering, Bangalore, hereby declare that, this assignment work entitled "PREDICTION OF DEMENTIA USING MRI DATASET" has been carried out by us under the guidance of **NAGARATHNA N,** Associate Professor, Department of CSE, BMS College of Engineering, Bangalore during the academic semester Aug-Dec 2017. We also declare that to the best of our knowledge and belief, the assignment reported here is not from part of any other report by any other students.

Signature of the Candidates

ANITHA D M (1BM14CS012)

.

AISHWARYA B (1BM14CS018)

ASHIKA M (1BM14CS023)

CONTENT **PAGE NO**

1. Introduction 1
2. Design 2
3. Implementation 4
4. Result 7
5. Conclusion 16
6. References 16

INTRODUCTION

Our study is based on MRI scans from publicly available database of Open Access Series of Imaging Studies (OASIS) which provides brain imaging data for analysis. The given data set consists of a longitudinal collection of 150 subjects covering the adult life span aged 60 to 96 including individuals with early-stage Alzheimer's Disease (AD). The subjects are all right-handed and include both men (62) and women (88). 64 of the subjects were diagnosed with very mild to mild AD at the time of their initial scan. 13 of the subjects were initially diagnosed as non-demented but converted to a diagnosis of very mild to mild AD at the time of a subsequent scan.

Problem Statement:

- Plot a correlogram which depicts the correlation of variables of the given dataset and displays the results graphically.
- To obtain a regression tree to show the variation from demented to non demented state.
- Prediction of Dementia.
- To show the percentage of male and female demented individuals.
- Categorizing and determining the count and the average age of demented subjects.

Proposed Solution:

- Collecting data from any source (websites like Kaggle) or do a survey to collect the necessary data.
- Explore the data and filter the data accordingly such as removing unnecessary fields, null values or filtering the data regionally.
- Visualize and find the correlation between the variables.
- Construct a decision tree in order to predict if the subject is demented or non demented.
- Make predictions on the filtered dataset.

DESIGN

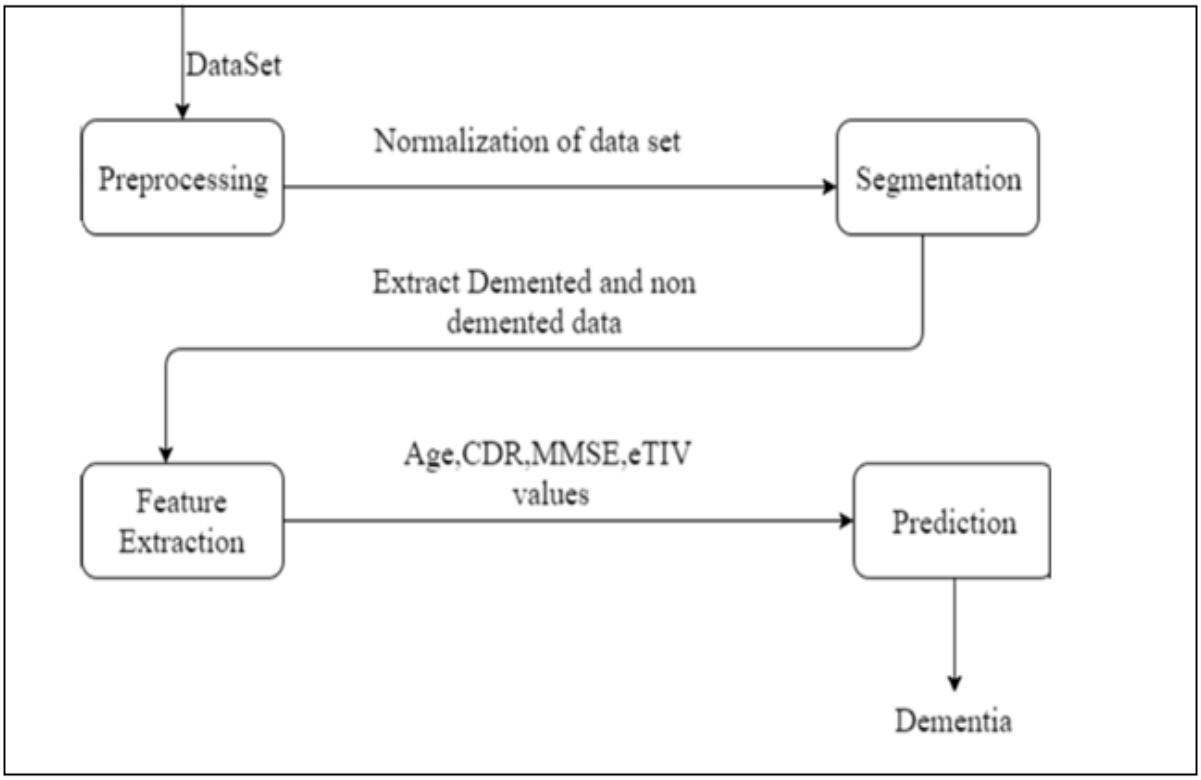

Figure 1 : The process followed in the analysis process.

- The dataset is obtained from the Open Access Series of Imaging Studies (OASIS) which provides brain imaging data for analysis.
- The obtained dataset is preprocessed and is normalized by removing the Null and unnecessary fields in the dataset.
- The data is then divided into demented and non-demented subsets.
- Based on our survey, we consider only certain parameters i.e. Age, CDR, MMSE, eTIV for analysis.
- We perform the analysis on the obtained data and predict dementia.

DATASET

dataset × dsr.R ×

Filter

	Subject.ID	MRI.ID	Group	Visit	MR.Delay	M.F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
1	OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R	87	14	2	27	0.0	1987	0.696	0.883
2	OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R	88	14	2	30	0.0	2004	0.681	0.876
6	OAS2_0004	OAS2_0004_MR1	Nondemented	1	0	F	R	88	18	3	28	0.0	1215	0.710	1.444
7	OAS2_0004	OAS2_0004_MR2	Nondemented	2	538	F	R	90	18	3	27	0.0	1200	0.718	1.462
8	OAS2_0005	OAS2_0005_MR1	Nondemented	1	0	M	R	80	12	4	28	0.0	1689	0.712	1.039
9	OAS2_0005	OAS2_0005_MR2	Nondemented	2	1010	M	R	83	12	4	29	0.5	1701	0.711	1.032
10	OAS2_0005	OAS2_0005_MR3	Nondemented	3	1603	M	R	85	12	4	30	0.0	1699	0.705	1.033
14	OAS2_0008	OAS2_0008_MR1	Nondemented	1	0	F	R	93	14	2	30	0.0	1272	0.698	1.380
15	OAS2_0008	OAS2_0008_MR2	Nondemented	2	742	F	R	95	14	2	29	0.0	1257	0.703	1.396
16	OAS2_0009	OAS2_0009_MR1	Demented	1	0	M	R	68	12	2	27	0.5	1457	0.806	1.205
17	OAS2_0009	OAS2_0009_MR2	Demented	2	576	M	R	69	12	2	24	0.5	1480	0.791	1.186
18	OAS2_0010	OAS2_0010_MR1	Demented	1	0	F	R	66	12	3	30	0.5	1447	0.769	1.213
19	OAS2_0010	OAS2_0010_MR2	Demented	2	854	F	R	68	12	3	29	0.5	1482	0.752	1.184
20	OAS2_0012	OAS2_0012_MR1	Nondemented	1	0	F	R	78	16	2	29	0.0	1333	0.748	1.316

Showing 1 to 15 of 354 entries

Figure 2 : A screenshot of the dataset with all parameters.

Attributes of the dataset:

- **Subject.ID:** This is used to identify any individual uniquely.
- **MRI.ID:** This identifies the scan number for every individual.
- **Group:** Group parameter is used to classify a person as either Demented, Non Demented or Converted.
- **Visit:** It is used to indicate the number of visits of the individual for the MRI scan.
- **MR.Delay:** The average latency in procuring the results of Magnetic Resonance Imaging.
- **M.F:** This tells the gender of the person.
- **Hand:** It shows the handedness of a person.
- **Age:** Age gives us information as to how old a person is.
- **EDUC:** The education completion level is indicated categorically in values between 6 to 23 (6 being the least educated and 23 being the most).
- **SES:** This tells us the socioeconomic status of an individual that ranges from 1 to 5.
- **MMSE:** The Mini–Mental State Examination or Folstein test is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment.
- **CDR:** The Clinical Dementia Rating or CDR is a numeric scale used to quantify the severity of symptoms of dementia (i.e. its stage).
- **eTIV:** The Estimated Total Intracranial Volume (eTIV) is an important covariate for volumetric analyses of the brain and brain regions.
- **nWBV:** It is the normalized Whole Brain Volume.
- **ASF:** It is the atlas scaling factor for the MRI scan.

IMPLEMENTATION

```
options(max.print=10000)
dataset<-read.csv("G:/Projects/DSR Project/oasis_longitudinal.csv")
dataset
View(dataset)
dataset<-na.omit(dataset)

#density plots for variables

attach(dataset)

plot(density(dataset$EDUC), main="Kernel Density of Education")
polygon(density(dataset$EDUC), col="darkblue", border="red")

plot(density(dataset$SES), main="Kernel Density of SES")
polygon(density(dataset$SES), col="darkblue", border="red")

plot(density(dataset$MMSE), main="Kernel Density of MMSE")
polygon(density(dataset$MMSE), col="darkblue", border="red")

plot(density(dataset$eTIV), main="Kernel Density of eTIV")
polygon(density(dataset$eTIV), col="darkblue", border="red")

plot(density(dataset$nWBV), main="Kernel Density of nWBV")
polygon(density(dataset$nWBV), col="darkblue", border="red")

#scatterplot

pairs(~CDR+nWBV+eTIV+MMSE,data = dataset,main="Scatterplot Matrix",
      col=dataset$Group,upper.panel=NULL)

#corrgram
install.packages("corrgram")
install.packages("dplyr")
library(corrgram)
library(dplyr)
library(rattle)
attach(dataset)
gsdd <-group_by(tbl_df(dataset))

corrgram(gsdd, lower.panel=panel.cor,
         upper.panel=panel.pie, text.panel=panel.txt,
         main="Correlation of interesting variables")
```

```
#heat tree

install.packages("rpart")
install.packages("rpart.plot")
library(rpart.plot)
library(rpart)
heat.tree <- function(tree, low.is.green=FALSE, ...) { # dots args passed to prp
  y <- tree$frame$yval
  if(low.is.green)
    y <- -y
  max <- max(y)
  min <- min(y)
  cols <- rainbow(99, end=.36)[
    ifelse(y > y[1], (y-y[1]) * (99-50) / (max-y[1]) + 50,
           (y-min) * (50-1) / (y[1]-min) + 1)]
  prp(tree, branch.col=cols, box.col=cols, ...)
}

heat.tree(reg_tree_1, type=4,cex=0.5, varlen=0, faclen=0, fallen.leaves=TRUE)

#prediction of dementia

fitTree<-rpart(Group~MMSE+CDR+eTIV+nWBV,dataset)
rpart.plot(fitTree,type=4,extra=2,clip.right.labs=FALSE,varlen=0,faclen=0)
newdata<-data.frame(MMSE=c(5,15,23,30),
                    CDR=c(0,0.5,1,2.0),
                    eTIV=c(1200,1500,1700,2000),
                    nWBV=c(0.6,0.7,0.75,0.8))
newdata
predict(fitTree,newdata,type="class")

#male-female percentage
install.packages("plotrix")
library(plotrix)

demented_data<-subset(dataset,Group=="Demented")
View(demented_data)
count_demented<-sum(complete.cases(demented_data))
count_demented

male_demented_data<-subset(demented_data,M.F=="M")
View(male_demented_data)
count_demented_male<-sum(complete.cases(male_demented_data))
count_demented_male
```

```
female_demented_data<-subset(demented_data,M.F=="F")
View(female_demented_data)
count_demented_female<-sum(complete.cases(female_demented_data))
count_demented_female

slices<-c(count_demented_male,count_demented_female)
percent<-round(slices/sum(slices)*100)
lbls<-c("Male","Female")
lbls<-paste(lbls, percent)
lbls<-paste(lbls,"%",sep="")
pie3D(slices,labels=lbls,explode=0.1,main="Pie Chart of Demented Individuals ")

#count and mean of demented individuals

interval <- seq(0, 100, by = 10)

sample<-cut(demented_data$Age,interval)

tapply(demented_data$Age, sample, sum)
count<-table(cut(demented_data$Age,interval))
count
average<-tapply(demented_data$Age, sample, mean)
average

bargraph<-plot(sample,main="Mean and Count of Demented Individuals",xlab="Age",ylab="Number of People",ylim=c(0,100),col="cyan")
points(x = bargraph, y = average,col="red",pch=19)
lines(x = bargraph, y = average,col="red")

#alzheimer's percentage

my.data.frame <- subset(demented_data, Age > 60 & MMSE < 27 & eTIV < 1600)
my.data.frame
count1<-sum(complete.cases(my.data.frame))
count1
count2<-sum(complete.cases(dataset))
count2
totalcount<-(count1/count2)*100
sprintf("The percentage of people suffering from Alzheimer's is %f percent.",totalcount)
```

RESULTS

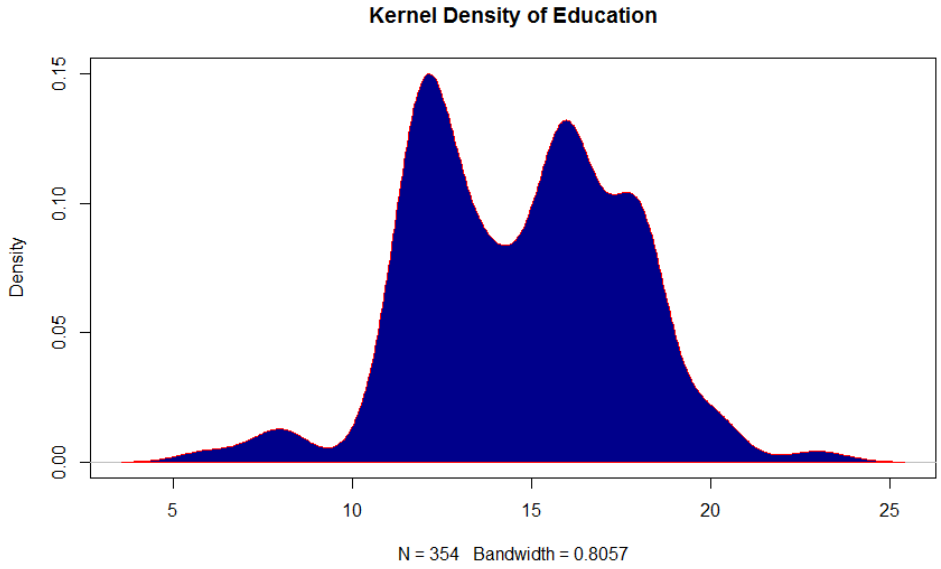

Figure 3.a : Probability density of Education

Education levels of most of the subjects lie in the range 10-19.

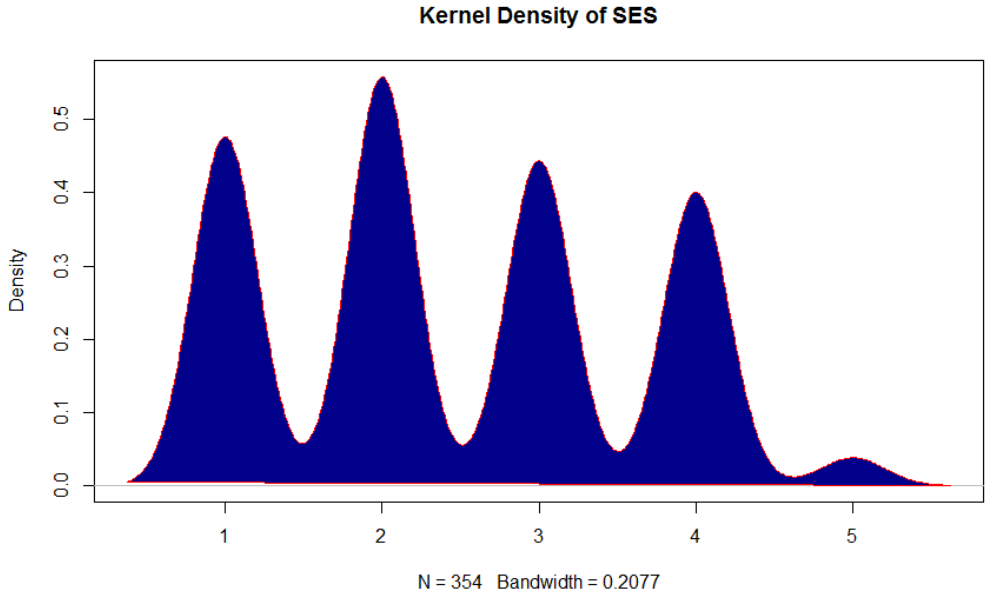

Figure 3.b : Probability density of SES

Most of the subjects have SES of 1 to 4, very few have the value of 5.

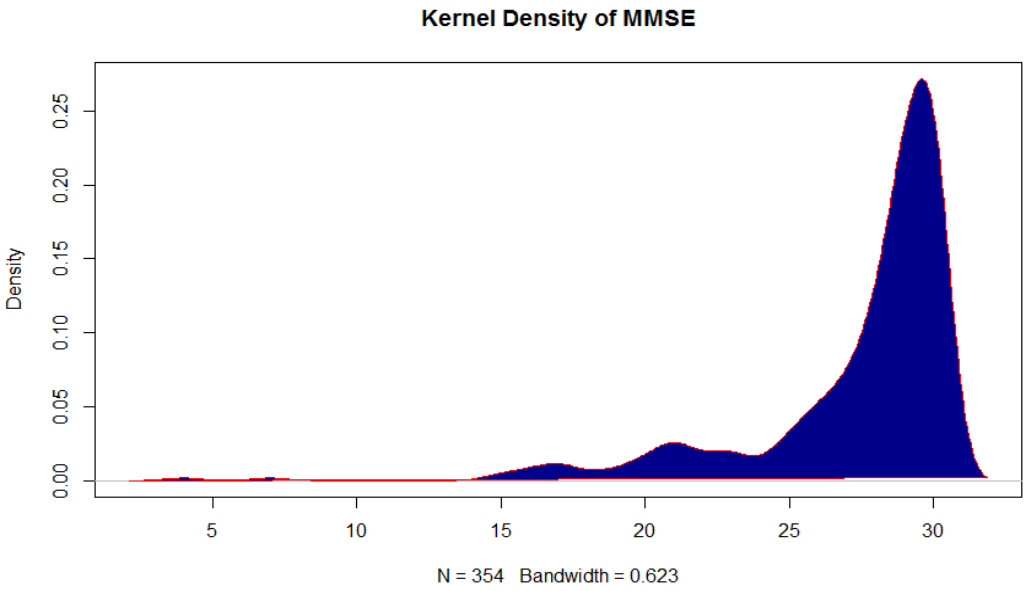

Figure 3.c : Probability density of MMSE

Majority of the subjects have MMSE values greater than 25, very few or almost none with values lower than 15.

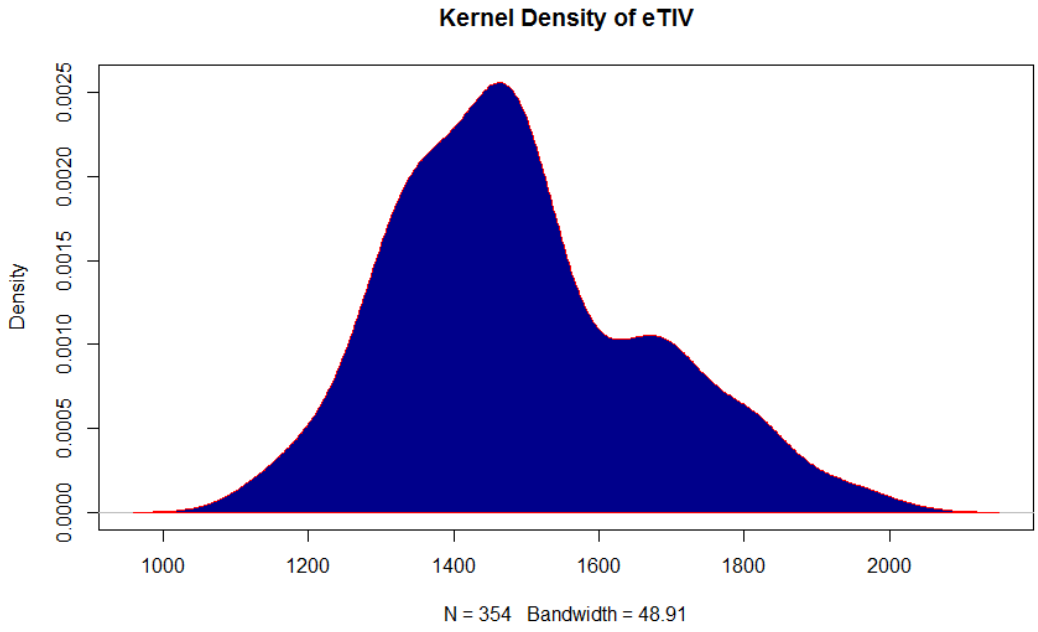

Figure 3.d : Probability density of eTIV

The eTIV values mostly range from 1300-1600mm^3.

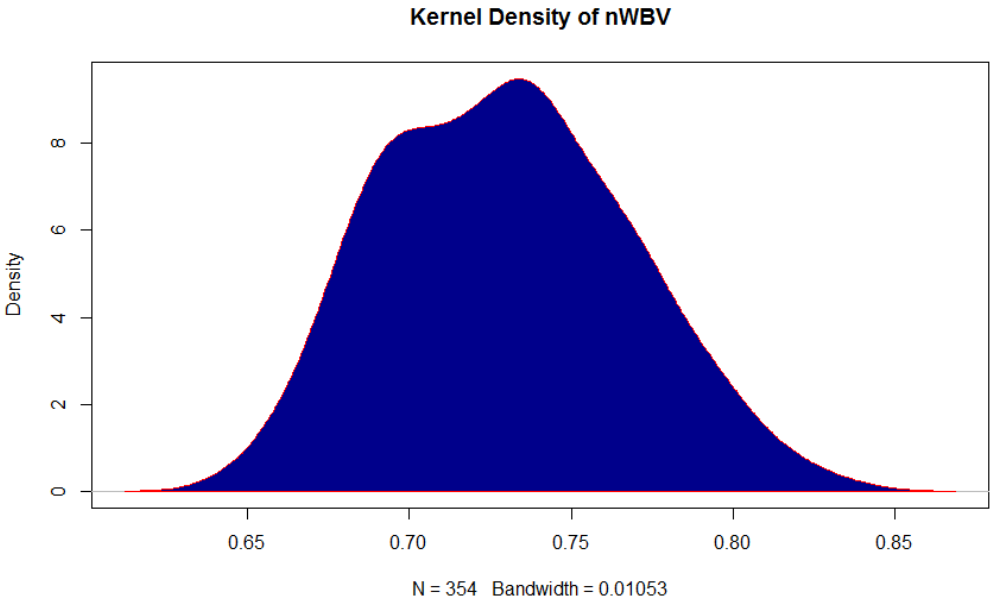

Figure 3.e : Probability density of nWBV

nWBV values are ranging from 0.65 to 0.85 with the peak at 0.74.

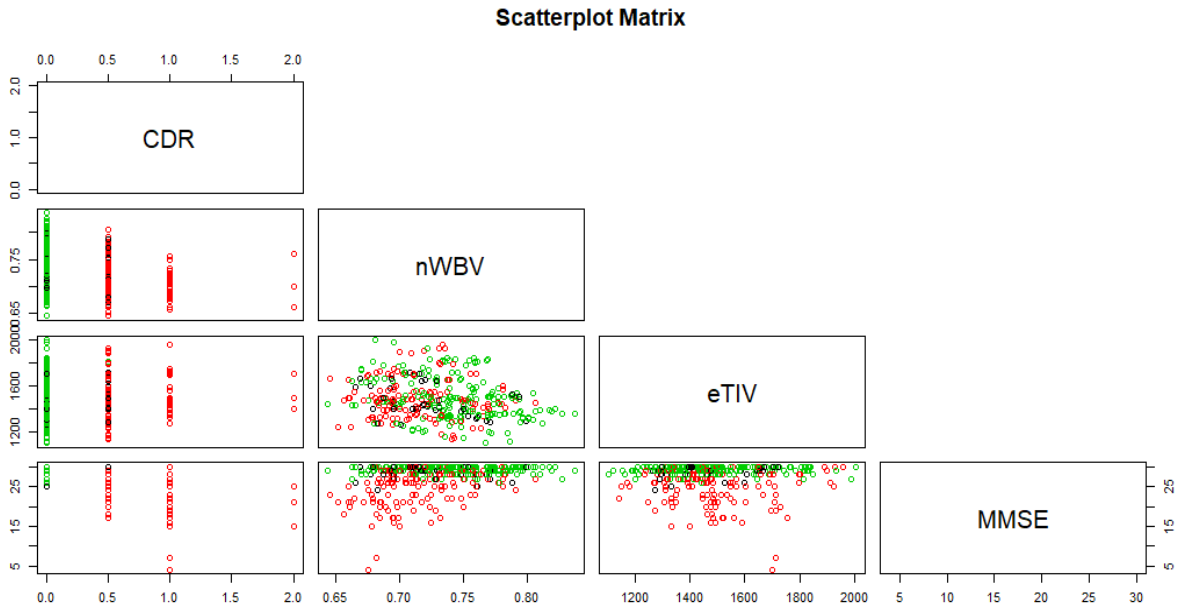

Figure 4 : Scatterplot depicting the relation between CDR, nWBV, eTIV and MMSE.

This figure shows the scatter plot of each parameter in relation with other parameters. The green dots indicate the Non Demented individuals, red dots indicate the demented individuals and black dots show the Converted individuals. From the above figure we can conclude that the MMSE values for Non Demented individuals is 25 or more.

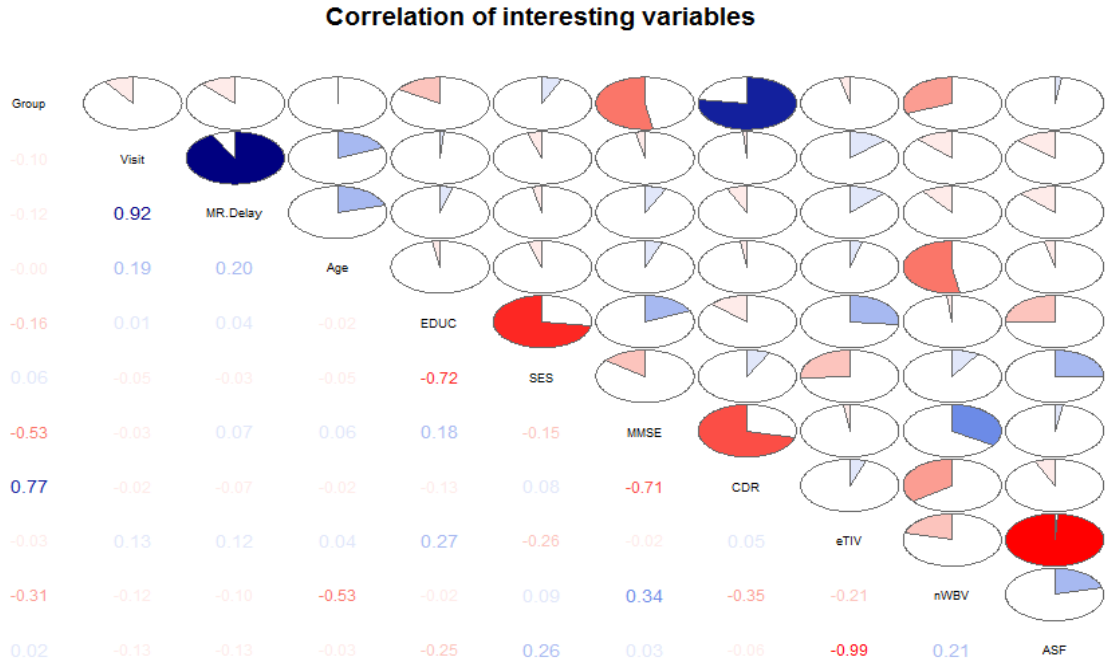

Figure 3 : Correlogram showing the interrelation between all the variables.

We can observe that the most related pairs of parameters are Group and Visit, Group and CDR. CDR which is an important parameter to predict Dementia is not determined by a single factor. MMSE and nWBV have a negative effect on CDR. eTIV has a very minimal effect on values of CDR.

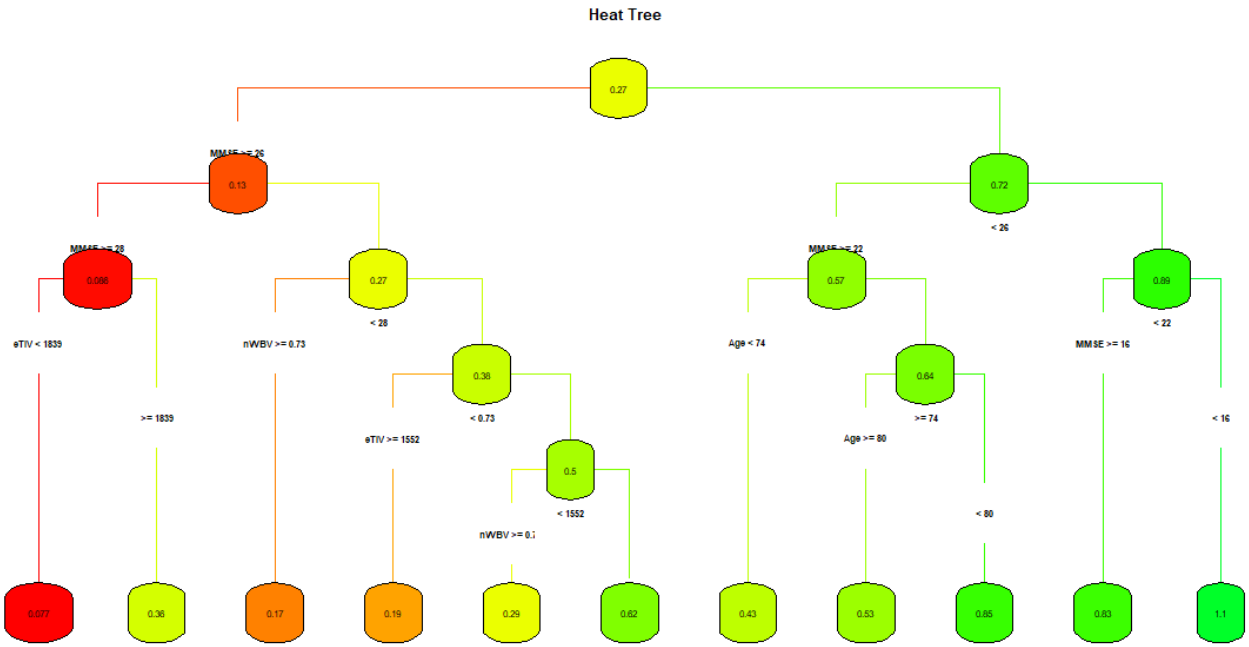

Figure 4 : Heat Tree depicting the variation from being non demented to demented.

The above graph depicts the Visualization of the given data set where the individuals are categorized as demented and non-demented. The darker shades of red color indicates possibility of being a non-demented individual and the darker shades of green color indicates the possibility of being a demented individual.

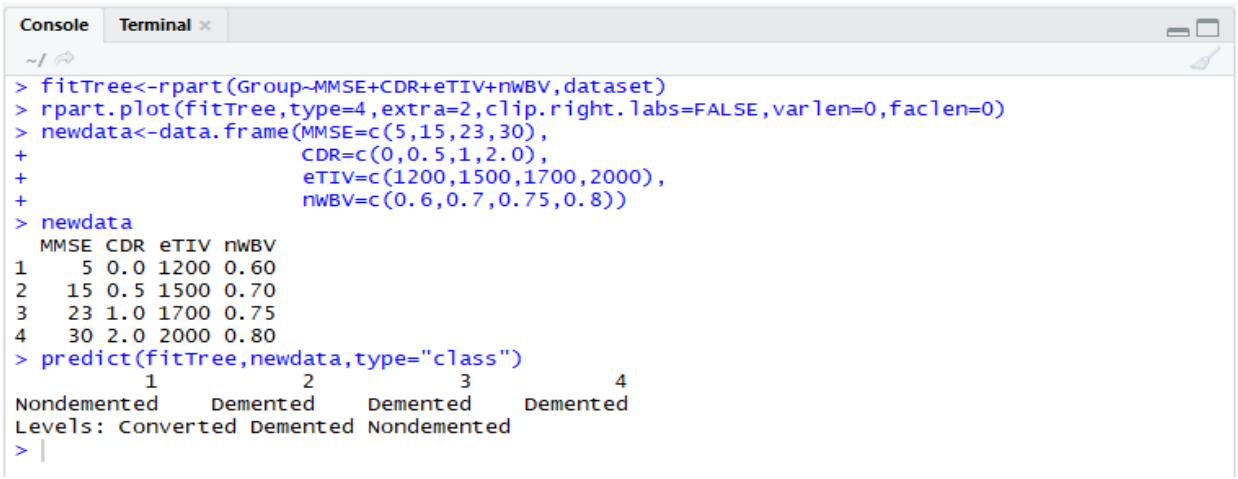

Figure 5 : A screenshot of the prediction for some input data.

Decision tree is used to predict the categorical value of "Demented" or "Non Demented" for a set of input values. In this case, for a given set of MMSE, CDR, eTIV and nWBV values it is predicted whether the individual is demented or not.

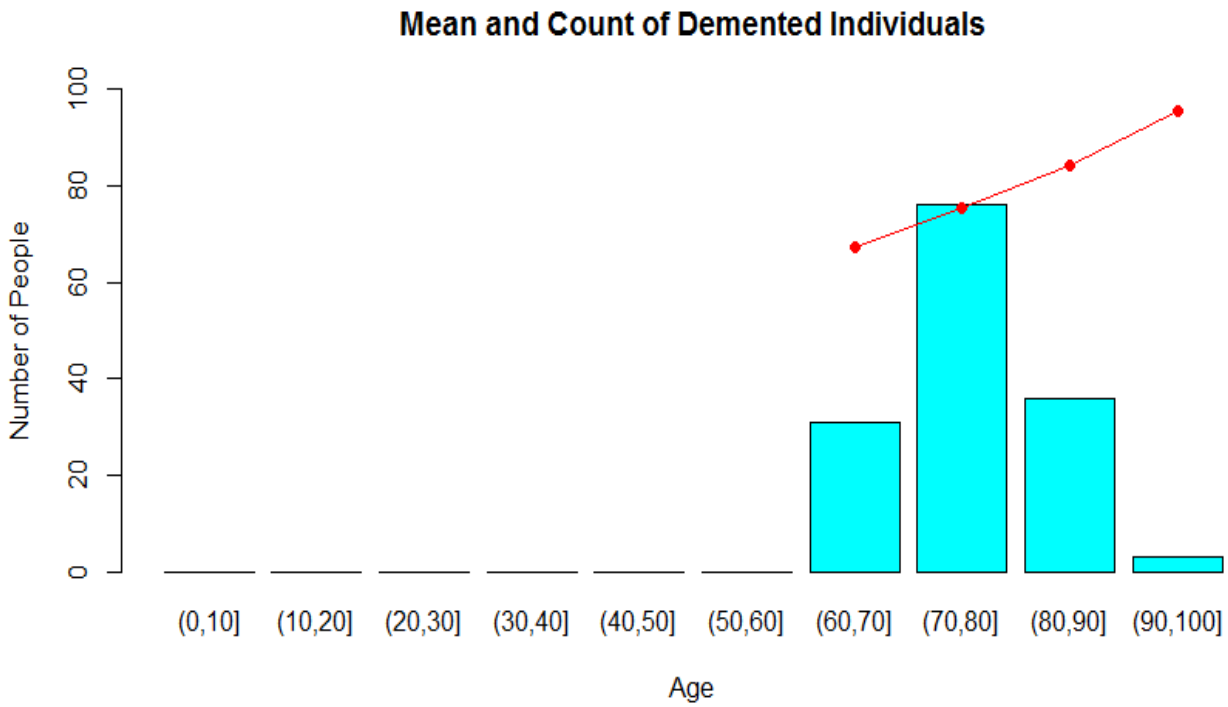

Figure 6 : A barplot showing the distribution and mean ages of individuals.

The above graph shows the mean and count of demented individuals.The bar plot represents the count of people in that particular interval.The line graph represents the average age of that interval. Our dataset mostly includes subjects in the age interval of 70-80.

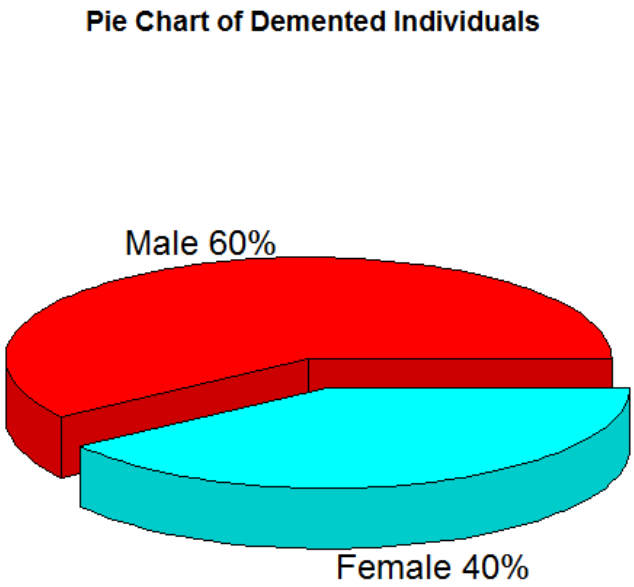

Figure 7 : Pie Chart depicting the male-female ratio of demented subjects

The above pie chart depicts the percentage of male and female individuals who are demented. The male ratio is greater (60%) for the given dataset.

CONCLUSION

From our analysis we can conclude that to determine if an individual is demented or not, CDR is the prime factor. CDR is affected by the following parameters:

- MMSE : CDR value increases with a decrease in the values of MMSE.
- eTIV : eTIV has a very minimal effect on CDR.
- nWBV : nWBV varies inversely with CDR.

Prediction of Dementia was done efficiently using the inbuilt functions of R and graph was plotted to show the demented and non demented individual.

REFERENCES

- https://pdfs.semanticscholar.org/7413/739f99c7e9225b14e8db68bf60101d56c77a.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2895005/
- https://www.hindawi.com/journals/cmmm/2014/862307/
- https://www.kaggle.com/abdulmagnun/compare-results-of-several-classifiers
- http://www.milbo.org/rpart-plot/prp.pdf