

# Final Project

Sindhu Bhattarai

6/12/2022

## Data Loading

```
train_genetic = read.csv(file="/Users/sakshyamdahal/Desktop/MS_Data_Science/Applied Predictive Modeling/Final project/train_genetic_disorders.csv", header= TRUE)
```

```
df <- train_genetic
#head(df,5)
dim(df)
```

```
> [1] 22083    45
```

## Initial feature reduction

Uninformative Feature Reduced

```
#drop unnecessary columns
drop_col <- c("Patient.Id", "Patient.First.Name", "Father.s.name", "Institute.Name", "Location.of.Institute", "Family.Name", "Parental.consent", "Status", "Place.of.birth", "Test.1", "Test.2", "Test.3", "Test.4", "Test.5")
df <- df[!(names(df) %in% drop_col)]
#names(df)
dim(df)
```

```
> [1] 22083    31
```

## Missing value in target vector.

```
#unique(df$Disorder.Subclass)
miss_value_target <- which(is.na(df$Disorder.Subclass) | (df$Disorder.Subclass == ""))
length(miss_value_target)
```

```
> [1] 3140
```

## Excuding data points where our target has missing value.

```
df <- df[-miss_value_target,]

dim(df)
```

```
> [1] 18943    31
```

## Check for class imbalance

```
prop.table(table(df$Disorder.Subclass))
```

```
>
>
>               Alzheimer's               Cancer
>           0.007812912           0.004856675
>           Cystic fibrosis           Diabetes
>           0.172992662           0.092012881
>           Hemochromatosis Leber's hereditary optic neuropathy
>           0.068151824           0.032043499
>           Leigh syndrome           Mitochondrial myopathy
>           0.258301219           0.221823365
>           Tay-Sachs
>           0.142004962
```

## Split the dataframe as per the disorder.subclass vector

There is certain class imbalance problem. We will split the data frame using `createDataPartition` function using `target class(Disorder.Subclass)` which helps us to get even balance of all classes in both train and test data

```
set.seed(100)
library(caret)
train_row <- createDataPartition(df$Disorder.Subclass, p = .80, list = FALSE)

train_df <- df[train_row,]
test_df <- df[-train_row,]

dim(df)
```

```
> [1] 18943    31
```

```
dim(train_df)
```

```
> [1] 15158    31
```

```
dim(test_df)
```

```
> [1] 3785    31
```

```
prop.table(table(train_df$Disorder.Subclass))
```

```
>
>               Alzheimer's               Cancer
>           0.007850640           0.004881911
>       Cystic fibrosis           Diabetes
>           0.172977965           0.092030611
> Hemochromatosis Leber's hereditary optic neuropathy
>           0.068148832           0.032062277
>       Leigh syndrome       Mitochondrial myopathy
>           0.258279456           0.221797071
>               Tay-Sachs
>           0.141971236
```

**\*\***We can see here no difference in the proportion of classes after split

```
#lapply(train_df, unique)
```

## Checking and handling missing values for train data

```
#changing ambiguous values from the data to na.
library(caret)
library(dplyr)
library(questionr)
train_df[train_df == ""] <- NA
test_df[test_df == ""] <- NA

train_df[train_df == "-"] <- NA
test_df[test_df == "-"] <- NA

#not changing not applicable to na as I think it itself is a class
df[df == "Not applicable"] <- NA

train_df[ train_df == "Not available"] <- NA
test_df[ test_df == "Not available"] <- NA

train_df[train_df == "No record"] <- NA
test_df[test_df == "No record"] <- NA

train_df[train_df == "None"] <- "No"
test_df[test_df == "None"] <- "No"

freq.na(train_df)
```

```

>
> Birth.asphyxia                missing %
> H.O.substance.abuse           4925 32
> H.O.radiation.exposure..x.ray. 4775 32
> Mother.s.age                 3962 26
> Father.s.age                 3898 26
> Maternal.gene                1824 12
> Symptom.2                    1458 10
> Respiratory.Rate..breaths.min. 1450 10
> Genetic.Disorder             1437  9
> Symptom.5                    1434  9
> Symptom.1                    1429  9
> Follow.up                    1428  9
> Birth.defects                1427  9
> White.Blood.cell.count..thousand.per.microliter. 1422  9
> Gender                       1407  9
> History.of.anomalies.in.previous.pregnancies 1389  9
> H.O.serious.maternal.illness 1388  9
> Blood.test.result            1388  9
> Symptom.4                    1382  9
> Heart.Rate..rates.min        1380  9
> No..of.previous.abortion     1378  9
> Folic.acid.details..peri.conceptional. 1376  9
> Symptom.3                    1376  9
> Assisted.conception.IVF.ART  1373  9
> Patient.Age                  971  6
> Autopsy.shows.birth.defect..if.applicable. 674  4
> Inherited.from.father        185  1
> Genes.in.mother.s.side        0  0
> Paternal.gene                 0  0
> Blood.cell.count..mcL.        0  0
> Disorder.Subclass            0  0

```

```
sum(is.na(train_df))
```

```
> [1] 54833
```

## Dropping the columns with high missing value

The imputaion in more than 50% of missing data might hamper our model.

```
#drop Birth.asphyxia column as 57% of data missing and imputation might create inaccurate model prediction
#dim(train_df)
train_df <- subset(train_df, select = -c(Birth.asphyxia))
dim(train_df)
```

```
> [1] 15158    30
```

```
test_df <- subset(test_df, select = -c(Birth.asphyxia))
dim(test_df)
```

```
> [1] 3785    30
```

```
str(train_df)
```

```
> 'data.frame': 15158 obs. of 30 variables:
> $ Patient.Age : int 2 4 6 12 11 14 3 11 4 6 .
..
> $ Genes.in.mother.s.side : chr "Yes" "Yes" "Yes" "Yes" .
..
> $ Inherited.from.father : chr "No" "Yes" "No" "No" ...
> $ Maternal.gene : chr "Yes" "No" "No" "Yes" ...
> $ Paternal.gene : chr "No" "No" "No" "No" ...
> $ Blood.cell.count..mcL. : num 4.76 4.91 4.89 4.71 4.72
...
> $ Mother.s.age : int NA NA 41 21 32 NA NA 45 4
4 NA ...
> $ Father.s.age : int NA 23 22 NA NA NA 63 44 4
2 NA ...
> $ Respiratory.Rate..breaths.min. : chr "Normal (30-60)" "Tachypnea" "Normal (30-60)" "Tachypnea" ...
ea" "Normal (30-60)" "Tachypnea" ...
> $ Heart.Rate..rates.min : chr "Normal" "Normal" "Tachycardia" "Normal" ...
ardia" "Normal" ...
> $ Follow.up : chr "High" "High" "Low" "High"
" ...
> $ Gender : chr NA NA NA "Male" ...
> $ Autopsy.shows.birth.defect..if.applicable. : chr "Not applicable" "No" "No"
t applicable" "No" ...
> $ Folic.acid.details..peri.conceptional. : chr "No" "Yes" "Yes" "No" ...
> $ H.O.serious.maternal.illness : chr NA "Yes" "No" "Yes" ...
> $ H.O.radiation.exposure...x.ray. : chr "No" "Not applicable" "Yes"
s" NA ...
> $ H.O.substance.abuse : chr "No" "Not applicable" NA
```

```

"Not applicable" ...
> $ Assisted.conception.IVF.ART : chr "No" "No" "Yes" NA ...
> $ History.of.anomalies.in.previous.pregnancies : chr "Yes" "Yes" "Yes" "Yes" .
..
> $ No..of.previous.abortion : int NA NA 4 1 4 0 3 0 1 1 ...
> $ Birth.defects : chr NA "Multiple" "Singular"
"Singular" ...
> $ White.Blood.cell.count..thousand.per.microliter.: num 9.86 5.52 NA 7.92 4.1 ...
> $ Blood.test.result : chr NA "normal" "normal" "inc
onclusive" ...
> $ Symptom.1 : int 1 1 0 0 0 1 0 1 0 1 ...
> $ Symptom.2 : int 1 NA 1 0 0 0 0 1 0 NA ...
> $ Symptom.3 : int 1 1 1 1 0 0 0 1 1 0 ...
> $ Symptom.4 : int 1 1 1 0 0 1 0 0 1 0 ...
> $ Symptom.5 : int 1 0 1 0 NA 0 0 1 1 NA ...
> $ Genetic.Disorder : chr "Mitochondrial genetic in
heritance disorders" NA "Multifactorial genetic inheritance disorders" "Mitochondrial
genetic inheritance disorders" ...
> $ Disorder.Subclass : chr "Leber's hereditary optic
neuropathy" "Cystic fibrosis" "Diabetes" "Leigh syndrome" ...

```

## Identifying the dtypes and changing into suitable dtypes.

1. Numeric vectors are mostly acceptable by all models. There are 11 numeric vectors in the data.
2. Vectors in string/character needs to be converted into factor for running smooth model. 19 character vectors are converted into factors.

```

#subsetting numeric columns and character for handling missing values and
library(dplyr)
# Subset numeric columns with dplyr
train_numeric <- select_if(train_df, is.numeric)
test_numeric <- select_if(test_df, is.numeric)

# Subset categorical columns with dplyr
train_categorical <- select_if(train_df,is.character)
test_categorical <- select_if(test_df,is.character)

dim(train_numeric)

```

```
> [1] 15158    11
```

```
dim(train_categorical)
```

```
> [1] 15158    19
```

```
dim(test_numeric)
```

```
> [1] 3785    11
```

```
dim(test_categorical)
```

```
> [1] 3785    19
```

```
lapply(train_categorical, unique)
```

```
> $Genes.in.mother.s.side
> [1] "Yes" "No"
>
> $Inherited.from.father
> [1] "No"  "Yes" NA
>
> $Maternal.gene
> [1] "Yes" "No"  NA
>
> $Paternal.gene
> [1] "No"  "Yes"
>
> $Respiratory.Rate..breaths.min.
> [1] "Normal (30-60)" "Tachypnea"      NA
>
> $Heart.Rate..rates.min
> [1] "Normal"          "Tachycardia" NA
>
> $Follow.up
> [1] "High" "Low"  NA
>
> $Gender
> [1] NA          "Male"      "Female"    "Ambiguous"
>
> $Autopsy.shows.birth.defect..if.applicable.
> [1] "Not applicable" "No"          "Yes"        NA
>
> $Folic.acid.details..peri.conceptional.
> [1] "No"  "Yes" NA
>
```



```

> $H.O.serious.maternal.illness
> [1] NA      "Yes" "No"
>
> $H.O.radiation.exposure..x.ray.
> [1] "No"      "Not applicable" "Yes"      NA
>
> $H.O.substance.abuse
> [1] "No"      "Not applicable" NA      "Yes"
>
> $Assisted.conception.IVF.ART
> [1] "No"  "Yes" NA
>
> $History.of.anomalies.in.previous.pregnancies
> [1] "Yes" "No"  NA
>
> $Birth.defects
> [1] NA      "Multiple" "Singular"
>
> $Blood.test.result
> [1] NA      "normal"      "inconclusive"
> [4] "slightly abnormal" "abnormal"
>
> $Genetic.Disorder
> [1] "Mitochondrial genetic inheritance disorders"
> [2] NA
> [3] "Multifactorial genetic inheritance disorders"
> [4] "Single-gene inheritance diseases"
>
> $Disorder.Subclass
> [1] "Leber's hereditary optic neuropathy" "Cystic fibrosis"
> [3] "Diabetes"                          "Leigh syndrome"
> [5] "Cancer"                            "Tay-Sachs"
> [7] "Hemochromatosis"                  "Mitochondrial myopathy"
> [9] "Alzheimer's"

```

```

#convert all character columns to factor
df_factor_train <- as.data.frame(unclass(train_categorical), stringsAsFactors = TRUE)
df_factor_test  <- as.data.frame(unclass(test_categorical), stringsAsFactors = TRUE)

str(df_factor_train)

```

```

> 'data.frame': 15158 obs. of 19 variables:
> $ Genes.in.mother.s.side : Factor w/ 2 levels "No","Yes": 2 2
2 2 2 2 2 1 1 2 ...
> $ Inherited.from.father : Factor w/ 2 levels "No","Yes": 1 2
1 1 1 1 1 1 2 1 ...
> $ Maternal.gene : Factor w/ 2 levels "No","Yes": 2 1
1 2 NA 2 2 2 2 NA ...
> $ Paternal.gene : Factor w/ 2 levels "No","Yes": 1 1
1 1 2 1 2 1 2 1 ...
> $ Respiratory.Rate..breaths.min. : Factor w/ 2 levels "Normal (30-60)
",...: 1 2 1 2 2 NA 1 2 2 2 ...
> $ Heart.Rate..rates.min : Factor w/ 2 levels "Normal","Tachy
cardia": 1 1 2 1 2 1 NA 2 2 NA ...
> $ Follow.up : Factor w/ 2 levels "High","Low": 1
1 2 1 2 2 2 2 2 ...
> $ Gender : Factor w/ 3 levels "Ambiguous","Fe
male",...: NA NA NA 3 3 2 3 3 3 3 ...
> $ Autopsy.shows.birth.defect..if.applicable. : Factor w/ 3 levels "No","Not appli
cable",...: 2 1 2 1 2 1 2 2 2 1 ...
> $ Folic.acid.details..peri.conceptional. : Factor w/ 2 levels "No","Yes": 1 2
2 1 1 1 NA 2 2 1 ...
> $ H.O.serious.maternal.illness : Factor w/ 2 levels "No","Yes": NA
2 1 2 2 1 2 2 1 2 ...
> $ H.O.radiation.exposure..x.ray. : Factor w/ 3 levels "No","Not appli
cable",...: 1 2 3 NA NA 1 1 1 1 3 ...
> $ H.O.substance.abuse : Factor w/ 3 levels "No","Not appli
cable",...: 1 2 NA 2 2 1 2 1 1 2 ...
> $ Assisted.conception.IVF.ART : Factor w/ 2 levels "No","Yes": 1 1
2 NA 2 NA 2 1 2 2 ...
> $ History.of.anomalies.in.previous.pregnancies: Factor w/ 2 levels "No","Yes": 2 2
2 2 1 1 1 2 2 NA ...
> $ Birth.defects : Factor w/ 2 levels "Multiple","Sin
gular": NA 1 2 2 1 1 1 1 1 2 ...
> $ Blood.test.result : Factor w/ 4 levels "abnormal","inc
onclusive",...: NA 3 3 2 NA 3 3 4 1 1 ...
> $ Genetic.Disorder : Factor w/ 3 levels "Mitochondrial
genetic inheritance disorders",...: 1 NA 2 1 2 3 3 1 2 3 ...
> $ Disorder.Subclass : Factor w/ 9 levels "Alzheimer's",.
.: 6 3 4 7 2 3 9 7 4 5 ...

```

```

#merge the transformed dataframe with numeric vectors data
train_df_transformed= cbind(train_numeric,df_factor_train)
test_df_transformed= cbind(test_numeric,df_factor_test)

str(train_df_transformed)

```

```

> 'data.frame': 15158 obs. of 30 variables:
> $ Patient.Age : int 2 4 6 12 11 14 3 11 4 6 .
..
> $ Blood.cell.count..mcL. : num 4.76 4.91 4.89 4.71 4.72
...
> $ Mother.s.age : int NA NA 41 21 32 NA NA 45 4
4 NA ...
> $ Father.s.age : int NA 23 22 NA NA NA 63 44 4
2 NA ...
> $ No..of.previous.abortion : int NA NA 4 1 4 0 3 0 1 1 ...
> $ White.Blood.cell.count..thousand.per.microliter.: num 9.86 5.52 NA 7.92 4.1 ...
> $ Symptom.1 : int 1 1 0 0 0 1 0 1 0 1 ...
> $ Symptom.2 : int 1 NA 1 0 0 0 0 1 0 NA ...
> $ Symptom.3 : int 1 1 1 1 0 0 0 1 1 0 ...
> $ Symptom.4 : int 1 1 1 0 0 1 0 0 1 0 ...
> $ Symptom.5 : int 1 0 1 0 NA 0 0 1 1 NA ...
> $ Genes.in.mother.s.side : Factor w/ 2 levels "No","Yes":
2 2 2 2 2 2 2 1 1 2 ...
> $ Inherited.from.father : Factor w/ 2 levels "No","Yes":
1 2 1 1 1 1 1 1 2 1 ...
> $ Maternal.gene : Factor w/ 2 levels "No","Yes":
2 1 1 2 NA 2 2 2 2 NA ...
> $ Paternal.gene : Factor w/ 2 levels "No","Yes":
1 1 1 1 2 1 2 1 2 1 ...
> $ Respiratory.Rate..breaths.min. : Factor w/ 2 levels "Normal (30
-60)",...: 1 2 1 2 2 NA 1 2 2 2 ...
> $ Heart.Rate..rates.min : Factor w/ 2 levels "Normal","T
achycardia": 1 1 2 1 2 1 NA 2 2 NA ...
> $ Follow.up : Factor w/ 2 levels "High","Low
": 1 1 2 1 2 2 2 2 2 2 ...
> $ Gender : Factor w/ 3 levels "Ambiguous"
,"Female",...: NA NA NA 3 3 2 3 3 3 3 ...
> $ Autopsy.shows.birth.defect..if.applicable. : Factor w/ 3 levels "No","Not a
pplicable",...: 2 1 2 1 2 1 2 2 2 1 ...
> $ Folic.acid.details..peri.conceptional. : Factor w/ 2 levels "No","Yes":
1 2 2 1 1 1 NA 2 2 1 ...
> $ H.O.serious.maternal.illness : Factor w/ 2 levels "No","Yes":
NA 2 1 2 2 1 2 2 1 2 ...
> $ H.O.radiation.exposure..x.ray. : Factor w/ 3 levels "No","Not a
pplicable",...: 1 2 3 NA NA 1 1 1 1 3 ...
> $ H.O.substance.abuse : Factor w/ 3 levels "No","Not a
pplicable",...: 1 2 NA 2 2 1 2 1 1 2 ...
> $ Assisted.conception.IVF.ART : Factor w/ 2 levels "No","Yes":
1 1 2 NA 2 NA 2 1 2 2 ...
> $ History.of.anomalies.in.previous.pregnancies : Factor w/ 2 levels "No","Yes":
2 2 2 2 1 1 1 2 2 NA ...
> $ Birth.defects : Factor w/ 2 levels "Multiple",

```

```
"Singular": NA 1 2 2 1 1 1 1 1 2 ...
> $ Blood.test.result : Factor w/ 4 levels "abnormal",
"inconclusive",...: NA 3 3 2 NA 3 3 4 1 1 ...
> $ Genetic.Disorder : Factor w/ 3 levels "Mitochondr
ial genetic inheritance disorders",...: 1 NA 2 1 2 3 3 1 2 3 ...
> $ Disorder.Subclass : Factor w/ 9 levels "Alzheimer'
s",...: 6 3 4 7 2 3 9 7 4 5 ...
```

**Mice imputation: MICE is a multiple imputation method used to replace missing data values in a data set under certain assumptions about the data missingness mechanism (e.g., the data are missing at random, the data are missing completely at random).**

```
# mice imputation
library(mice)
set.seed(100)
train_df_imputed <- mice(train_df_transformed, m=5, maxit = 3, method = 'pmm')
```

```
>
> iter imp variable
> 1 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 2 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 4 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
```

```

lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 5 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 2 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion* White.
Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.
4 Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min.*
Heart.Rate..rates.min Follow.up* Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies* Birth.defects Blood.test.result Genetic.Disorder
> 2 2 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter.* Symptom.1 Symptom.2 Symptom.3 Symptom.
4 Symptom.5 Inherited.from.father Maternal.gene* Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender* Autopsy.shows.birth.defect..if.applicable.
* Folic.acid.details..peri.conceptional.* H.O.serious.maternal.illness H.O.radiati
on.exposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART History.of.an
omalies.in.previous.pregnancies* Birth.defects Blood.test.result Genetic.Disorder
> 2 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1* Symptom.2* Symptom.3 Symptom
.4 Symptom.5 Inherited.from.father* Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender* Autopsy.shows.birth.defect..if.applicable.
* Folic.acid.details..peri.conceptional.* H.O.serious.maternal.illness H.O.radiati
on.exposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.ano
malies.in.previous.pregnancies* Birth.defects Blood.test.result Genetic.Disorder
> 2 4 Patient.Age* Mother.s.age Father.s.age No..of.previous.abortion White.
Blood.cell.count..thousand.per.microliter.* Symptom.1 Symptom.2 Symptom.3 Symptom
.4 Symptom.5 Inherited.from.father Maternal.gene* Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable.*
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies* Birth.defects Blood.test.result* Genetic.Disorder
> 2 5 Patient.Age* Mother.s.age Father.s.age No..of.previous.abortion* White
.Blood.cell.count..thousand.per.microliter.* Symptom.1 Symptom.2* Symptom.3 Sympt
om.4 Symptom.5* Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min
. Heart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicabl
e. Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiati
on.exposure..x.ray.* H.O.substance.abuse* Assisted.conception.IVF.ART* History.of.

```

```

anomalies.in.previous.pregnancies Birth.defects Blood.test.result* Genetic.Disorde
r
> 3 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up* Gender* Autopsy.shows.birth.defect..if.applicable. F
olic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.ex
posure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalie
s.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 3 2 Patient.Age* Mother.s.age Father.s.age No..of.previous.abortion* White
.Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom
.4 Symptom.5 Inherited.from.father Maternal.gene* Respiratory.Rate..breaths.min.*
Heart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 3 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3* Symptom.
4 Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. H
eart.Rate..rates.min Follow.up* Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness* H.O.radiation.
exposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART History.of.anoma
lies.in.previous.pregnancies Birth.defects Blood.test.result* Genetic.Disorder
> 3 4 Patient.Age* Mother.s.age Father.s.age* No..of.previous.abortion White
.Blood.cell.count..thousand.per.microliter.* Symptom.1* Symptom.2 Symptom.3 Sympt
om.4 Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min.
* Heart.Rate..rates.min Follow.up Gender* Autopsy.shows.birth.defect..if.applicab
le. Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiat
ion.exposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART* History.of.a
nomalies.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 3 5 Patient.Age Mother.s.age Father.s.age* No..of.previous.abortion White.
Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.
4 Symptom.5 Inherited.from.father* Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable.*
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART* History.of.anoma
lies.in.previous.pregnancies Birth.defects Blood.test.result* Genetic.Disorder

```

```
test_df_imputed <- mice(test_df_transformed, m=5, maxit = 3, method = 'pmm')
```

```

>
> iter imp variable
> 1 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol

```

```

ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 2 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 4 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 1 5 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.expo
sure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies.
in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 2 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1* Symptom.2 Symptom.3* Symptom
.4 Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up* Gender Autopsy.shows.birth.defect..if.applicable.
* Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiatio
n.exposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART History.of.ano
malies.in.previous.pregnancies* Birth.defects Blood.test.result* Genetic.Disorder
> 2 2 Patient.Age* Mother.s.age Father.s.age No..of.previous.abortion White.
Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2* Symptom.3 Symptom
.4* Symptom.5* Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min* Follow.up Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness* H.O.radiation.
exposure..x.ray.* H.O.substance.abuse Assisted.conception.IVF.ART History.of.ano
malies.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 2 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter.* Symptom.1 Symptom.2* Symptom.3 Symptom

```

```
.4 Symptom.5 Inherited.from.father* Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 2 4 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min.* Hea
rt.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fo
lic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.exp
osure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies
.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 2 5 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion* White.
Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.
4* Symptom.5 Inherited.from.father Maternal.gene* Respiratory.Rate..breaths.min.
Heart.Rate..rates.min* Follow.up* Gender Autopsy.shows.birth.defect..if.applicable
. Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiatio
n.exposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART History.of.ano
malies.in.previous.pregnancies Birth.defects Blood.test.result Genetic.Disorder
> 3 1 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter.* Symptom.1 Symptom.2 Symptom.3 Symptom.
4 Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. H
eart.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies* Birth.defects Blood.test.result Genetic.Disorder
> 3 2 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up* Gender Autopsy.shows.birth.defect..if.applicable. Fo
lic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.exp
osure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies
.in.previous.pregnancies* Birth.defects Blood.test.result Genetic.Disorder
> 3 3 Patient.Age Mother.s.age Father.s.age No..of.previous.abortion White.B
lood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom.4
Symptom.5 Inherited.from.father Maternal.gene Respiratory.Rate..breaths.min. Hear
t.Rate..rates.min Follow.up Gender Autopsy.shows.birth.defect..if.applicable. Fol
ic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.exp
osure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomalies
.in.previous.pregnancies* Birth.defects Blood.test.result* Genetic.Disorder
> 3 4 Patient.Age* Mother.s.age Father.s.age No..of.previous.abortion* White
.Blood.cell.count..thousand.per.microliter. Symptom.1 Symptom.2 Symptom.3 Symptom
.4 Symptom.5* Inherited.from.father* Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min* Follow.up* Gender Autopsy.shows.birth.defect..if.applicable
. Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiatio
n.exposure..x.ray. H.O.substance.abuse* Assisted.conception.IVF.ART History.of.ano
malies.in.previous.pregnancies Birth.defects Blood.test.result* Genetic.Disorder
```



```
> 3 5 Patient.Age Mother.s.age* Father.s.age No..of.previous.abortion White.
Blood.cell.count..thousand.per.microliter. Symptom.1* Symptom.2* Symptom.3 Sympto
m.4 Symptom.5 Inherited.from.father* Maternal.gene Respiratory.Rate..breaths.min.
Heart.Rate..rates.min Follow.up Gender* Autopsy.shows.birth.defect..if.applicable.
Folic.acid.details..peri.conceptional. H.O.serious.maternal.illness H.O.radiation.e
xposure..x.ray. H.O.substance.abuse Assisted.conception.IVF.ART History.of.anomali
es.in.previous.pregnancies* Birth.defects Blood.test.result* Genetic.Disorder
```

```
set.seed(100)
train_df_imputed <- complete(train_df_imputed,3)
test_df_imputed <- complete(test_df_imputed,3)

sum(is.na(train_df_imputed))
```

```
> [1] 0
```

```
sum(is.na(test_df_imputed))
```

```
> [1] 0
```

## EDA