**IMPORTING DATA:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

**DATA CLEANING:**

**----Data cleaning**

**----Removing the nulls**

DELETE FROM dbo.obesity\_data

WHERE Age IS NULL

OR gender IS NULL

OR Veg\_intake IS NULL

OR No\_of\_meals\_intake IS NULL

OR Water\_intake IS NULL

or Physcial\_Activity is null

or Height is null

or weight is null ---(o rows affected)

**---Removing decimals from columns:**

update dbo.Obesity\_data SET Age = Round(Age,0) ----(2111 rows affected)

update dbo.Obesity\_data SET Veg\_intake = Round(Veg\_intake,0) ----(2111 rows affected)

update dbo.Obesity\_data SET No\_of\_meals\_intake = Round(No\_of\_meals\_intake,0) ----(2111 rows affected)

update dbo.Obesity\_data SET Water\_intake = round(Water\_intake,0) ----(2111 rows affected)

update dbo.Obesity\_data SET Physcial\_Activity = round (Physcial\_Activity,0) ----(2111 rows affected)

update dbo.Obesity\_data SET Tech\_using\_time= round (Tech\_using\_time, 0) ----(2111 rows affected)

ALTER TABLE dbo.obesity\_data ALTER COLUMN Tech\_using\_time decimal(9,0)

ALTER TABLE dbo.obesity\_data ALTER column Height decimal(9,2)

ALTER TABLE dbo.obesity\_data ALTER COLUMN Weight decimal (9,0)

Select Age,Height,Weight,Veg\_intake,No\_of\_meals\_intake,Water\_intake, Physcial\_Activity,Tech\_using\_time from dbo.Obesity\_data

A screenshot of a computer

Description automatically generated

**-----BINARIZATION CODE:**

----Family history with over weight

Select count (family\_history\_with\_overweight) as freq\_no from dbo.Obesity\_data where family\_history\_with\_overweight = 'no'---385

Select count (family\_history\_with\_overweight) as freq\_no from dbo.Obesity\_data where family\_history\_with\_overweight = 'yes'--1726

UPDATE dbo.Obesity\_data SET family\_history\_with\_overweight = REPLACE(family\_history\_with\_overweight, 'no','0')

Select count (family\_history\_with\_overweight) as freq\_no from dbo.Obesity\_data where family\_history\_with\_overweight = '0' ---385

UPDATE dbo.Obesity\_data SET family\_history\_with\_overweight = REPLACE(family\_history\_with\_overweight, 'yes','1')

Select count (family\_history\_with\_overweight) as freq\_no from dbo.Obesity\_data where family\_history\_with\_overweight = '1'---1726

----high cal intake:

Select count (High\_Cal\_intake) as freq\_no from dbo.Obesity\_data where High\_Cal\_intake = 'no'---245

Select count (High\_Cal\_intake) as freq\_no from dbo.Obesity\_data where High\_Cal\_intake = 'yes'--1866

UPDATE dbo.Obesity\_data SET High\_Cal\_intake = REPLACE(High\_Cal\_intake, 'no','0')

Select count (High\_Cal\_intake) as freq\_no from dbo.Obesity\_data where High\_Cal\_intake = '0' ---245

UPDATE dbo.Obesity\_data SET High\_Cal\_intake = REPLACE(High\_Cal\_intake, 'yes','1')

Select count (High\_Cal\_intake) as freq\_no from dbo.Obesity\_data where High\_Cal\_intake = '1'----1866

-----Consu\_food\_intake\_betw\_Meals

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = 'no'---51

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = 'sometimes'--1765

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = 'Always'---53

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = 'Frequently'---242

UPDATE dbo.Obesity\_data SET Consu\_food\_intake\_betw\_Meals = REPLACE(Consu\_food\_intake\_betw\_Meals, 'no','0')

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = '0' ---51

UPDATE dbo.Obesity\_data SET Consu\_food\_intake\_betw\_Meals = REPLACE(Consu\_food\_intake\_betw\_Meals, 'Sometimes','1')

UPDATE dbo.Obesity\_data SET Consu\_food\_intake\_betw\_Meals = REPLACE(Consu\_food\_intake\_betw\_Meals, 'Always','1')

UPDATE dbo.Obesity\_data SET Consu\_food\_intake\_betw\_Meals = REPLACE(Consu\_food\_intake\_betw\_Meals, 'Frequently','1')

Select count (Consu\_food\_intake\_betw\_Meals) as freq\_no from dbo.Obesity\_data where Consu\_food\_intake\_betw\_Meals = '1'---2060

-----Smoke

Select count (SMOKE) as freq\_no from dbo.Obesity\_data where SMOKE = 'no'---2067

Select count (SMOKE) as freq\_no from dbo.Obesity\_data where SMOKE = 'yes'--44

UPDATE dbo.Obesity\_data SET SMOKE = REPLACE(SMOKE, 'no','0')

Select count (SMOKE) as freq\_no from dbo.Obesity\_data where SMOKE = '0' ---2067

UPDATE dbo.Obesity\_data SET SMOKE = REPLACE(SMOKE, 'yes','1')

Select count (SMOKE) as freq\_no from dbo.Obesity\_data where SMOKE = '1'----44

-----cal\_intake

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring = 'no'

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring = 'yes'

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring= 'sometimes'--1401

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring= 'Always'---1

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring = 'Frequently'----70

UPDATE dbo.Obesity\_data SET Cal\_intake\_Monitoring = REPLACE(Cal\_intake\_Monitoring, 'no','0')

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring = '0'

UPDATE dbo.Obesity\_data SET Cal\_intake\_Monitoring= REPLACE(Cal\_intake\_Monitoring, 'yes','1')

UPDATE dbo.Obesity\_data SET Cal\_intake\_Monitoring = REPLACE(Cal\_intake\_Monitoring, 'Sometimes','1')

UPDATE dbo.Obesity\_data SET Cal\_intake\_Monitoring = REPLACE(Cal\_intake\_Monitoring, 'Always','1')

UPDATE dbo.Obesity\_data SET Cal\_intake\_Monitoring = REPLACE(Cal\_intake\_Monitoring, 'Frequently','1')

Select count (Cal\_intake\_Monitoring) as freq\_no from dbo.Obesity\_data where Cal\_intake\_Monitoring = '1'---1472

----Alcohol

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = 'no'---639

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = 'sometimes'--1401

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = 'Always'---1

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = 'Frequently'---70

UPDATE dbo.Obesity\_data SET Alcohol\_intake = REPLACE(Alcohol\_intake, 'no','0')

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = '0' ---639

UPDATE dbo.Obesity\_data SET Alcohol\_intake = REPLACE(Alcohol\_intake, 'Sometimes','1')

UPDATE dbo.Obesity\_data SET Alcohol\_intake = REPLACE(Alcohol\_intake, 'Always','1')

UPDATE dbo.Obesity\_data SET Alcohol\_intake = REPLACE(Alcohol\_intake, 'Frequently','1')

Select count (Alcohol\_intake) as freq\_no from dbo.Obesity\_data where Alcohol\_intake = '1'---1472

-----water intake

Select count (Water\_intake) as freq\_no from dbo.Obesity\_data where Water\_intake = '1'---485

Select count (Water\_intake) as freq\_no from dbo.Obesity\_data where Water\_intake = '2'--1110

Select count (Water\_intake) as freq\_no from dbo.Obesity\_data where Water\_intake = '3'---516

UPDATE dbo.Obesity\_data SET Water\_intake = REPLACE(Water\_intake, '1','0')

Select count (Water\_intake) as freq\_no from dbo.Obesity\_data where Water\_intake = '0' ---485

UPDATE dbo.Obesity\_data SET Water\_intake = REPLACE(Water\_intake, '2','1')

UPDATE dbo.Obesity\_data SET Water\_intake = REPLACE(Water\_intake, '3','1')

Select count (Water\_intake) as freq\_no from dbo.Obesity\_data where Water\_intake = '1'---1626

---Physical activity:

Select count (Physcial\_Activity) as freq\_no from dbo.Obesity\_data where Physcial\_Activity = '1'---776

Select count (Physcial\_Activity) as freq\_no from dbo.Obesity\_data where Physcial\_Activity = '2'--496

Select count (Physcial\_Activity) as freq\_no from dbo.Obesity\_data where Physcial\_Activity = '3'---119

UPDATE dbo.Obesity\_data SET Physcial\_Activity = REPLACE(Physcial\_Activity, '1','0')

Select count (Physcial\_Activity) as freq\_no from dbo.Obesity\_data where Physcial\_Activity = '0' ---1496

UPDATE dbo.Obesity\_data SET Physcial\_Activity = REPLACE(Physcial\_Activity, '2','1')

UPDATE dbo.Obesity\_data SET Physcial\_Activity = REPLACE(Physcial\_Activity, '3','1')

Select count (Physcial\_Activity) as freq\_no from dbo.Obesity\_data where Physcial\_Activity = '1'---615

----Veg intake

Select count (Veg\_intake) as freq\_no from dbo.Obesity\_data where Veg\_intake= '1'---102

Select count (Veg\_intake) as freq\_no from dbo.Obesity\_data where Veg\_intake = '2'--1013

Select count (Veg\_intake) as freq\_no from dbo.Obesity\_data where Veg\_intake = '3'---996

UPDATE dbo.Obesity\_data SET Veg\_intake = REPLACE(Veg\_intake, '1','0')

Select count (Veg\_intake) as freq\_no from dbo.Obesity\_data where Veg\_intake = '0' ---102

UPDATE dbo.Obesity\_data SET Veg\_intake = REPLACE(Veg\_intake, '2','1')

UPDATE dbo.Obesity\_data SET Veg\_intake = REPLACE(Veg\_intake, '3','1')

Select count (Veg\_intake) as freq\_no from dbo.Obesity\_data where Veg\_intake = '1'---2009

---no of meals intake:

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake = '1'---316

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake = '2'--176

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake = '3'---1470

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake = '4'---149

UPDATE dbo.Obesity\_data SET No\_of\_meals\_intake = REPLACE(No\_of\_meals\_intake, '1','0')

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake= '0' ---316

UPDATE dbo.Obesity\_data SET No\_of\_meals\_intake = REPLACE(No\_of\_meals\_intake, '2','1')

UPDATE dbo.Obesity\_data SET No\_of\_meals\_intake = REPLACE(No\_of\_meals\_intake, '3','1')

UPDATE dbo.Obesity\_data SET No\_of\_meals\_intake = REPLACE(No\_of\_meals\_intake, '4','1')

Select count (No\_of\_meals\_intake) as freq\_no from dbo.Obesity\_data where No\_of\_meals\_intake = '1'---1795

A screenshot of a calendar

Description automatically generated with medium confidence

**-----CALCULATE BMI**

ALTER TABLE dbo.obesity\_data ADD BMI AS round((weight/(Height\*Height)),2) persisted;

SELECT Height, Weight, FLOOR(BMI) AS BMI

FROM dbo.Obesity\_data;

Select \* from dbo.Obesity\_data

A screenshot of a computer

Description automatically generated with medium confidence

**-----CATEGORIZATION OF BMI ATTRIBUTE AS UNDERWEIGHT, NORMAL, OVERWEIGHT**

---adding obesity levels:

Select height, weight, BMI,

CASE

When BMI <=18 then 'underweight'

When BMI between 18 and 25 Then 'Normal'

When BMI >25 then 'Overweight'

END as obesity\_level into #BMI\_table From dbo.Obesity\_data ---2111 rows affected

Select \* from #BMI\_table

----drop unwanted columns from the table:

Alter table dbo.obesity\_data drop column Tech\_using\_time

Alter table dbo.obesity\_data drop column obesity\_level

Alter table dbo.obesity\_data drop column transportation\_used

**----Join the tables**

----Joining #BMI\_table and Obesity\_data

SELECT a.\*, b.obesity\_level

INTO dbo.final\_table1

FROM dbo.Obesity\_data a

JOIN #BMI\_table b ON b.BMI = a.BMI

GROUP BY a.Gender, a.Age, a.Height, a.Weight, a.family\_history\_with\_overweight,

a.High\_Cal\_intake, a.Veg\_intake, a.No\_of\_meals\_intake,

a.Consu\_food\_intake\_betw\_Meals, a.SMOKE, a.Alcohol\_intake,

a.Water\_intake, a.Cal\_intake\_Monitoring, a.Physcial\_Activity, a.BMI, b.obesity\_level

ORDER BY a.Gender, a.Age, a.Height, a.Weight, a.family\_history\_with\_overweight,

a.High\_Cal\_intake, a.Veg\_intake, a.No\_of\_meals\_intake,

a.Consu\_food\_intake\_betw\_Meals, a.SMOKE, a.Alcohol\_intake,

a.Water\_intake, a.Cal\_intake\_Monitoring, a.Physcial\_Activity, a.BMI, b.obesity\_level; ----1766 rows affected

Select \* from dbo.final\_table1

A screenshot of a calendar

Description automatically generated with medium confidence

**CODE FOR SPLITTING DATA**

select distinct Age

into #OBA from dbo.final\_table1---40 rows affected

select top 80 percent \*

into #train\_obA

from #OBA---32 rows affected

select \*

into [trainingOB]

from dbo.final\_table1

where Age in (select \* from #train\_obA) ---1512 rows affected

select \*

into [testingOB2]

from dbo.final\_table1

where Age not in (select \* from #train\_obA)---254 rows

**WEKA results:**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

**Logistic regression for training data set:**

A screenshot of a computer

Description automatically generated

**Random tree: Training data set:**

A screenshot of a computer

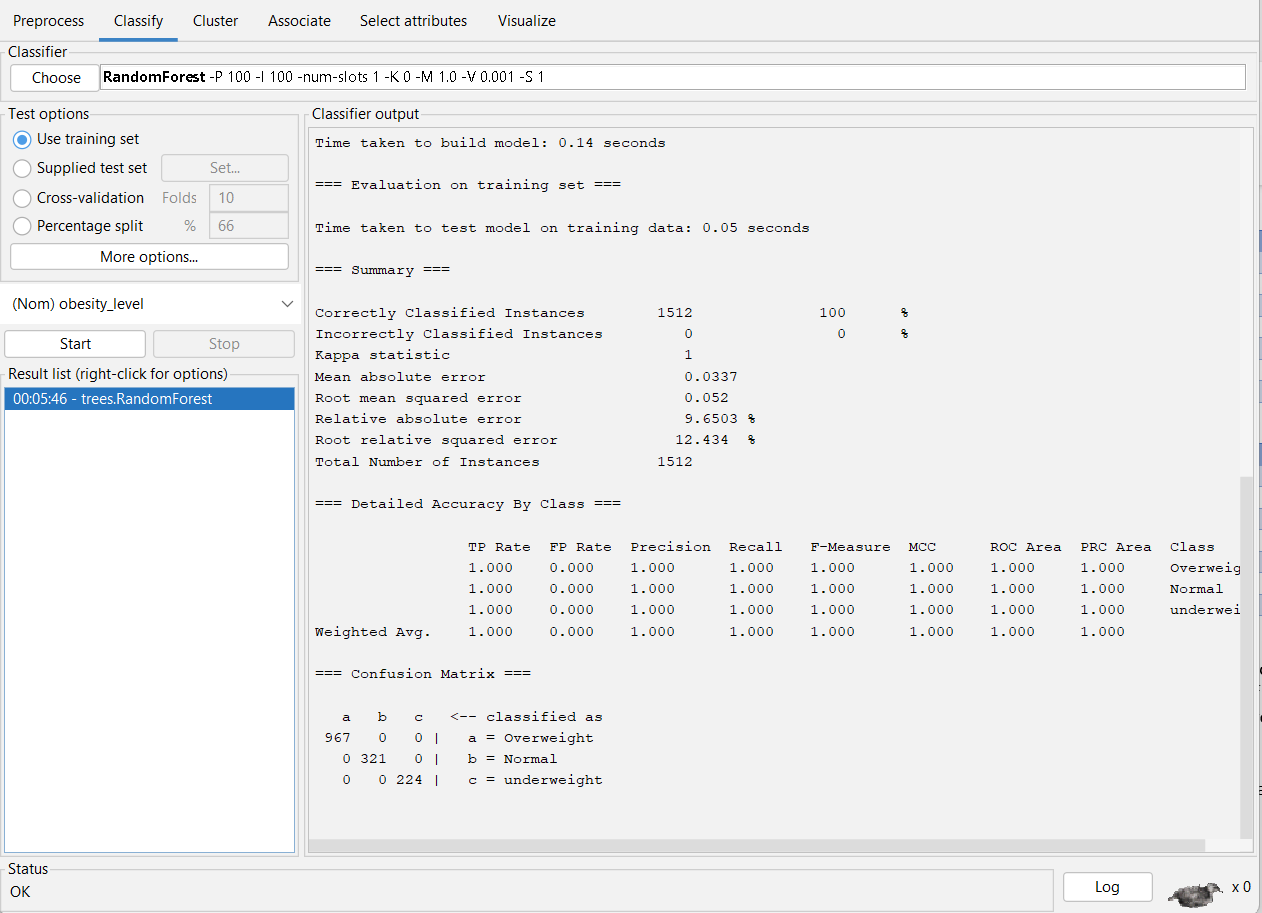
Description automatically generated

**Naïve bays- Training data set:**

A screenshot of a computer

Description automatically generated

**Random forest- Training data set:**



**Logistic regression for test data set:**

A screenshot of a computer

Description automatically generated

**Random tree :**

A screenshot of a computer

Description automatically generated

**Random forest:**

A screenshot of a computer

Description automatically generated

**Naïve bays:**

A screenshot of a computer

Description automatically generated

**FP growth:**

A picture containing text, screenshot, colorfulness

Description automatically generated