Date: _____ Day: M T W T F S S

M. Tanzeel Saleem
04071913026

IML - Assignment #3

Regression Tree (Karachi House Price Prediction)

→ Var (S) = 0.66

(i) Type :- (Flat, House)

Var (Flat) = 0.67

var (House) = 0.65

Gain = 0.0

(ii) Bedrooms :- (2, 3) :-

Var (Bedrooms-2) = 0.2

var (Bedrooms-3) = 0.66

Gain = 0.15

(iii) Bathroom :- (2, 3)

var (Bathroom-2) = 0.2

var (Bathroom-3) = 0.66

Gain = 0.15

(iv) Location (DHA, Bahria) :-

var (Bahria) = 0.58

Var (DHA) = 0.73

Gain = 0.02

(v) Area (1, 0) OR (≤236, >236)

var (≤236) = 0.43

var (>236) = 0

Gain = 0.26

MaX - Gain = 0.26 (Area)

Date: _____ Day: M T W T F S S

→ Var (S) = 0.43

(i) Type:

var (Flat) = 0.67

var (House) = 0.04

Gain = 0.03

(ii) Bedrooms:-

var (Bedrooms-2) = 0.2

var (Bedrooms-3) = 0.41

Gain = 0.09

(iii) Bathroom:-

var (Bathroom-2) = 0.2

var (Bathroom-3) = 0.41

Gain = 0.09

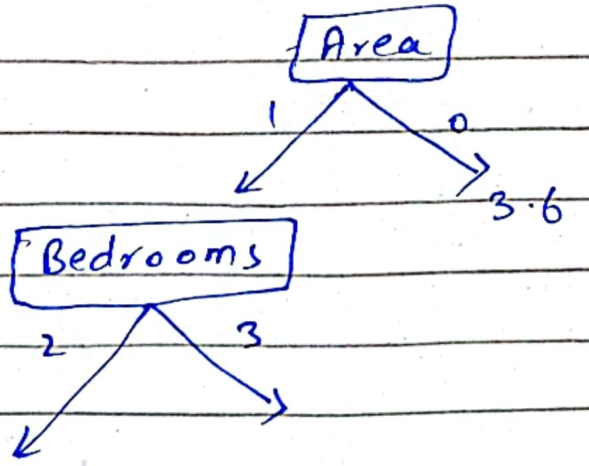
(iv) Location:-

Var (Bahria) = 0.09

var (DHA) = 0.73

Gain = 0.07

Max - Gain = 0.09 (Bedrooms, Bathroom)



Date: _____ Day: M T W T F S S

→ Var(S) = 0.2

(i) Type :-

var (Flat) = 0.0

var (House) = 0.0

Gain = 0.0

(ii) Bedrooms :-

var (Bedrooms-2) = 0.0

var (Bedrooms-3) = 0.0

Gain = 0.0

(iii) Bathroom :

var (Bathroom-2) = 0.0

var (Bathroom-3) = 0.0

Gain = 0.0

(iv) Location :-

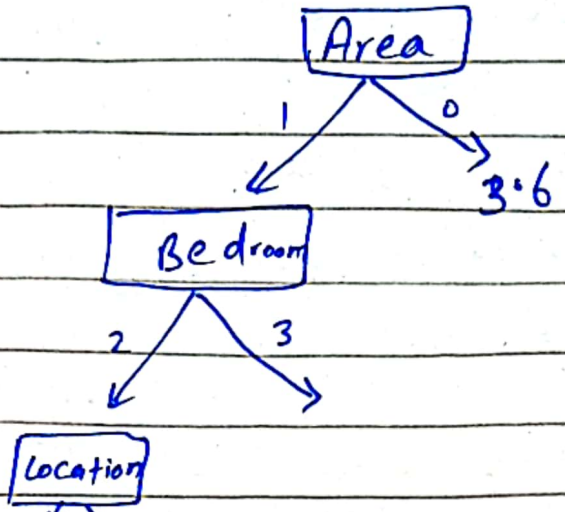
var (Bahria) = 0.23

var (DHA) = 0.15

Gain = 0.02

Max-Gain = 0.02 (Location)

Note: _____ Day: M T W T F S S



→ $Var^{B}(S) = 0.4$

(i) Type :-

Var (Flat) = 0.03

Var (House) = 0.04

Gain = 0.37

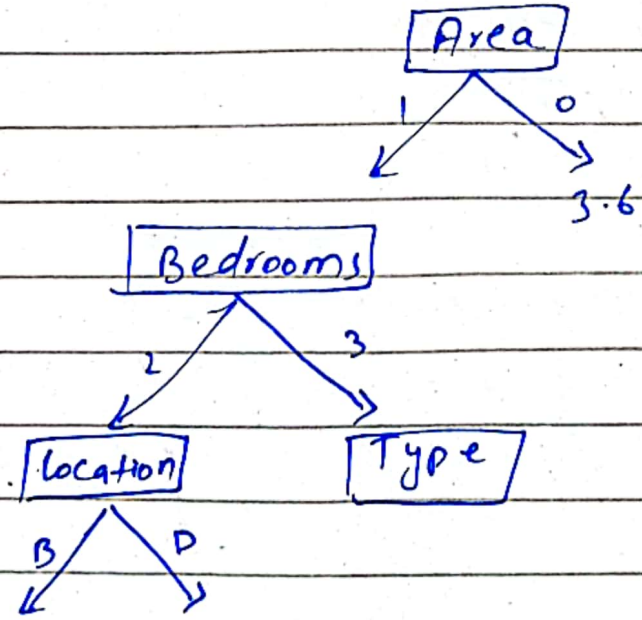
(ii) Bathroom :

Var (Bathroom-2) = 0.0

Var (Bathroom-3) = 0.0

Gain = 0.0

Max- Gain = 0.37 (TYPE)



Date: _____ Day: S M T W T F S S

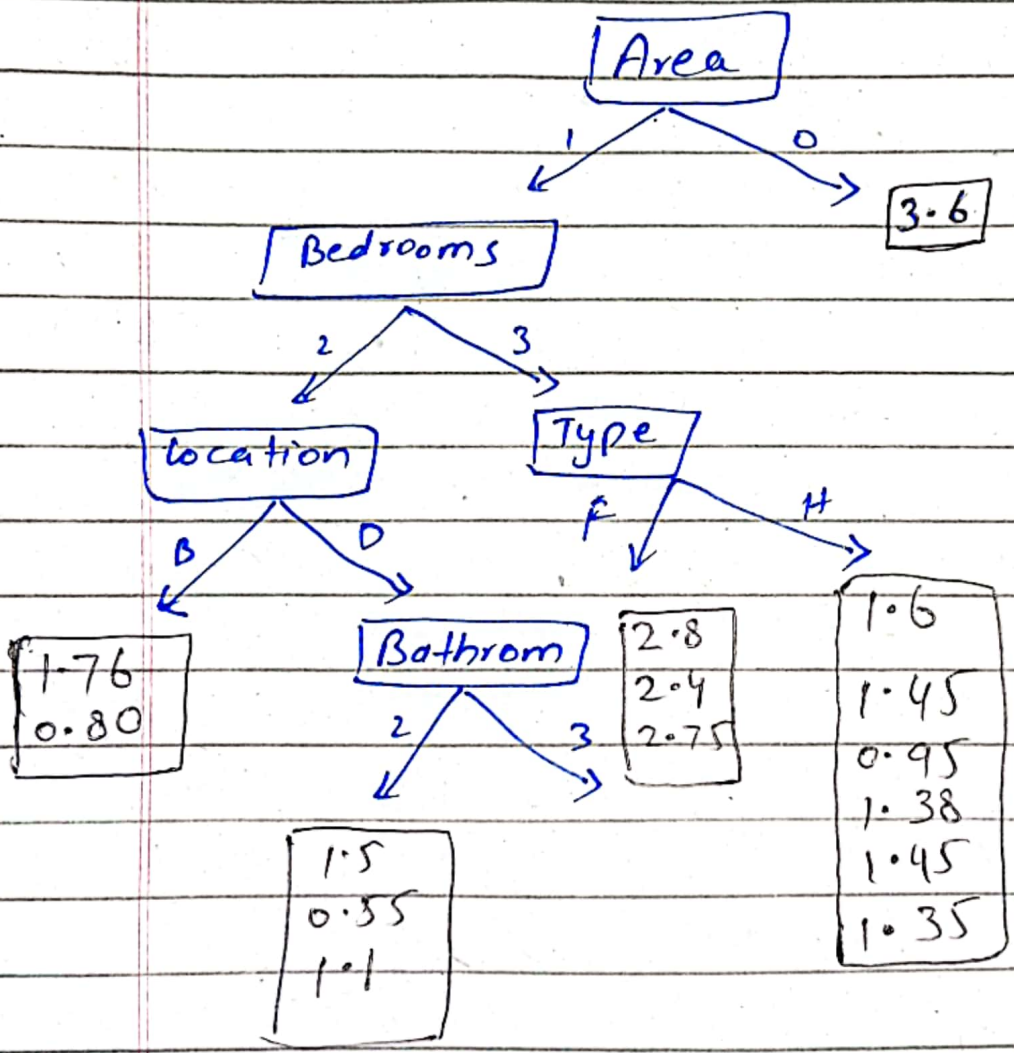
→ var (S) 0.23

(i) Bathroom:-

Var (Bathroom-2) = 0.15

Var (Bathroom-3) = 0.03

Gain = 0.64



Date: _____ Day: M T W T F S S

⇒ Random Forest Tree - I

→ Var(S) = 0.73558

(i) Bedrooms:

V(Bedrooms-2) = 0.3000

V(Bedrooms-3) = 0.6827

Gain = 0.1549

(ii) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0.6827

Gain = 0.1549

(iii) Type:

V(Flat) = 0.7400

V(House) = 0.7048

Gain = 0.0119

(iv) Location:-

V(Bahria) = 0.6186

V(DHA) = 0.8445

Gain = 0.01146

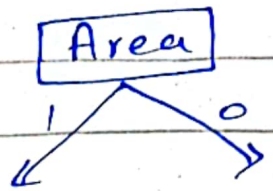
(v) Area:

V(≤236) = 0.5241

V(>236) = 0

Gain) = 0.2464

Max-Gain = 0.2464 (Area)



Date: _____ Day: M T W T F S S

→ Var(S) = 0.5241

(i) Bedrooms:

V(Bedrooms-2) = 0.3000

V(Bedrooms-3) = 0.4623

Gain = 0.1188

(ii) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0.4623

Gain = 0.1188

(iii) Type:

V(Flat) = 0.7400

V(House) = 0.0643

Gain = 0.0736

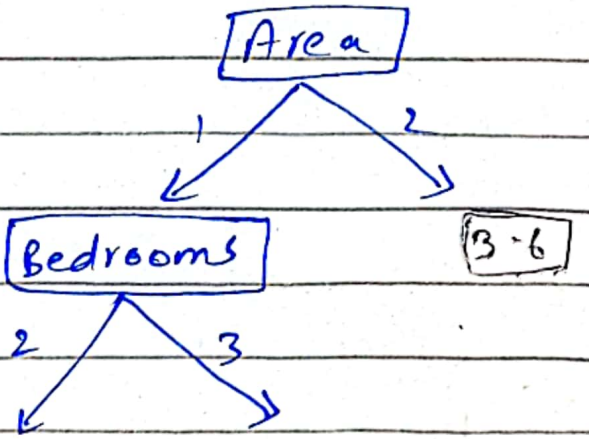
(iv) Location:-

V(Bahria) = 0.0814

V(DHA) = 0.8445

Gain = 0.0995

Max-Gain = 0.1188 (Bedrooms, Bathroom)



Date: _____ Day: M T W T F S S

→ Var (S) = 0.3000

(i) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0

Gain = 0

(ii) Type :-

V(Flat) = 0.3000

V(House) = 0

Gain = 0

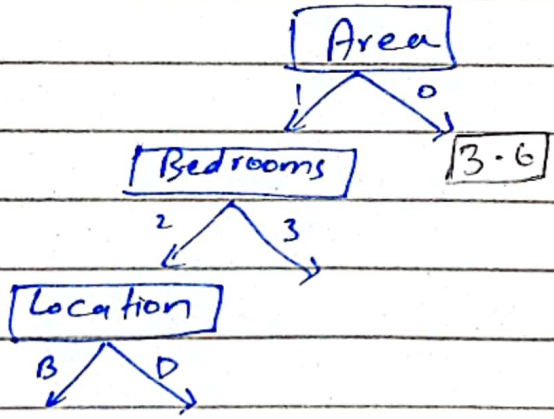
(iii) Location =

V(Bahria) = 0.2

V(DHA) = 0.2005

Gain = 0.1496

Max- Gain = 0.1496 (Location)



→ Var(S) = 0.4623

(i) Bathroom:

V(Bathroom-2) = 0

V(Bathroom-3) = 0.4623

Gain = 0

Date: _____ Day: M T W T F S S

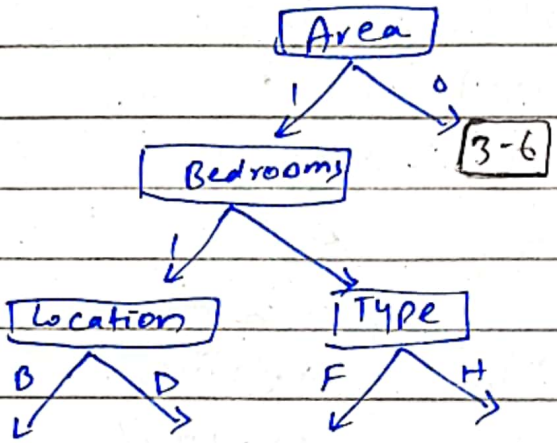
(ii) Type:

V(Flat) = 0.04

V(House) = 0.0643

Gain = 0.4076

Max-Gain = 0.4076 (Type)



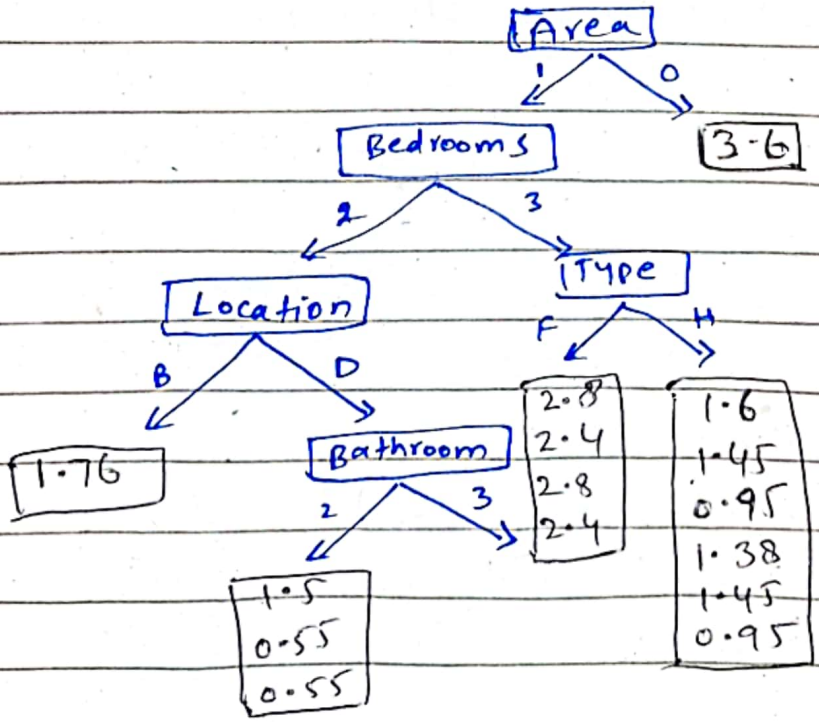
→ Var(S) = 0.08445

(i) Bathroom:

V(Bathroom-2) = 0.2005

V(Bathroom-3) = 0.04

Gain = 0.7357



Date: _____ Day: M T W T F S S

⇒ Random Forest Tree II :-

→ Var(S) = 0.8207

(i) Bedrooms:-

V(Bedrooms-2) = 0.2704

V(Bedrooms-3) = 0.8371

Gain = 0.01759

(ii) Bathroom:-

V(Bathroom-2) = 0.2704

V(Bathroom-3) = 0.8371

Gain = 0.01759

(iii) Type:-

V(Flat) = 0.6880

V(House) = 0.7898

Gain = 0.01969

(iv) Location:-

V(Bahria) = 0.841644

V(DHA) = 0.7584

Gain = 0.0668

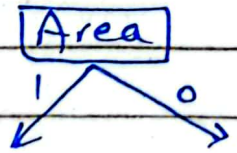
(v) Area:-

V(<236) = 0.6563

V(>236) = 0

Gain) = 0.0437

Max-Gain = 0.0437 (Area)



→ var(S) = 0.6563

(i) Bedrooms:

Date: _____ Day: M T W T F S S

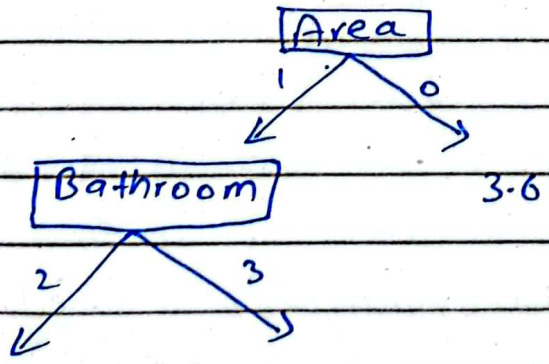
var(Bedrooms-2) = 0.2704
var(Bedroom-3) = 0.6745
Gain = 0.0320

(ii) Bathroom:-
v(Bathroom-2) = 0.2704
v(Bathroom-3) = 0.6745
Gain = 0.0320

(iii) Type:-
v(Flat) = 0.6406
v(House) = 0.6167
Gain = 0.02929

(iv) Location:-
v(Bahria) = 0.5501
v(DHA) = 0.7584
Gain = 0.03181

Max-Gain:- 0.0320 (Bathroom, Bedroom)



→ var(S) = 0.2704

(i) Bedrooms:-
var(Bedroom-2) = 0.270
var(Bedroom-3) = 0
Gain = 0

Date: _____ Day: M T W T F S S

(ii) Type:-

v(Flat) = 0.2704

v(House) = 0

Gain = 0

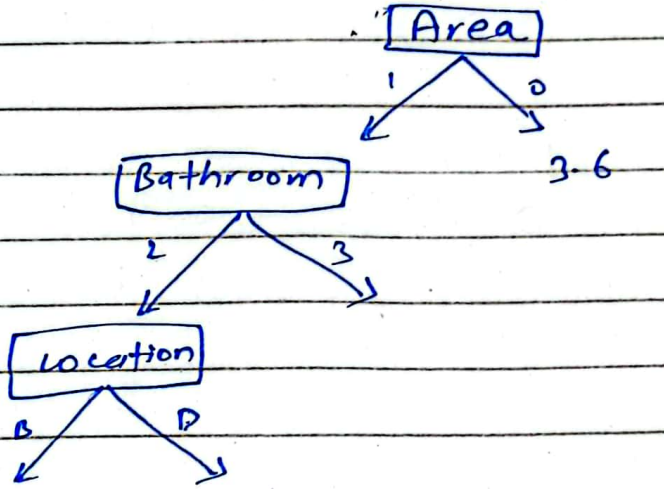
(iii) Location =

v(Bahria) = 0

v(DHA) = 0.2256

Gain = 0.1200

Max - Gain = 0.1200 (location)



3.6

→ var(S) = 0.5501

(i) Bedrooms:-

v(Bedroom-2) = 0

v(Bedroom-3) = 0.6167

Gain = 0.0018

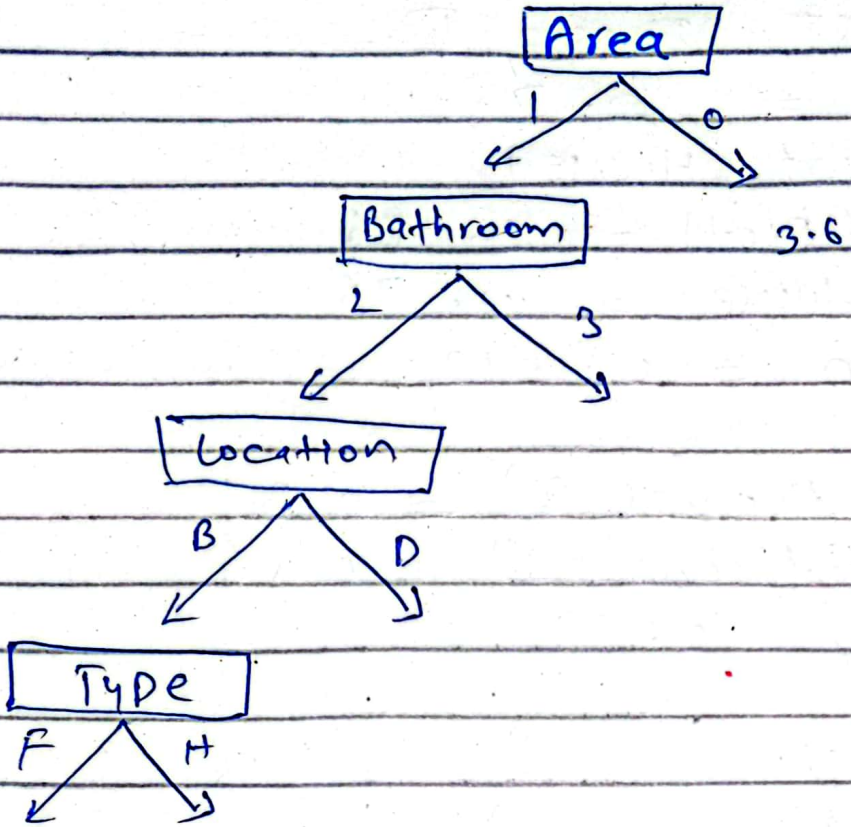
(ii) Type:-

v(Flat) = 0

v(House) = 0.6167

Gain = 0.0018

Max - Gain = 0.0018 (Bedrooms, Type)



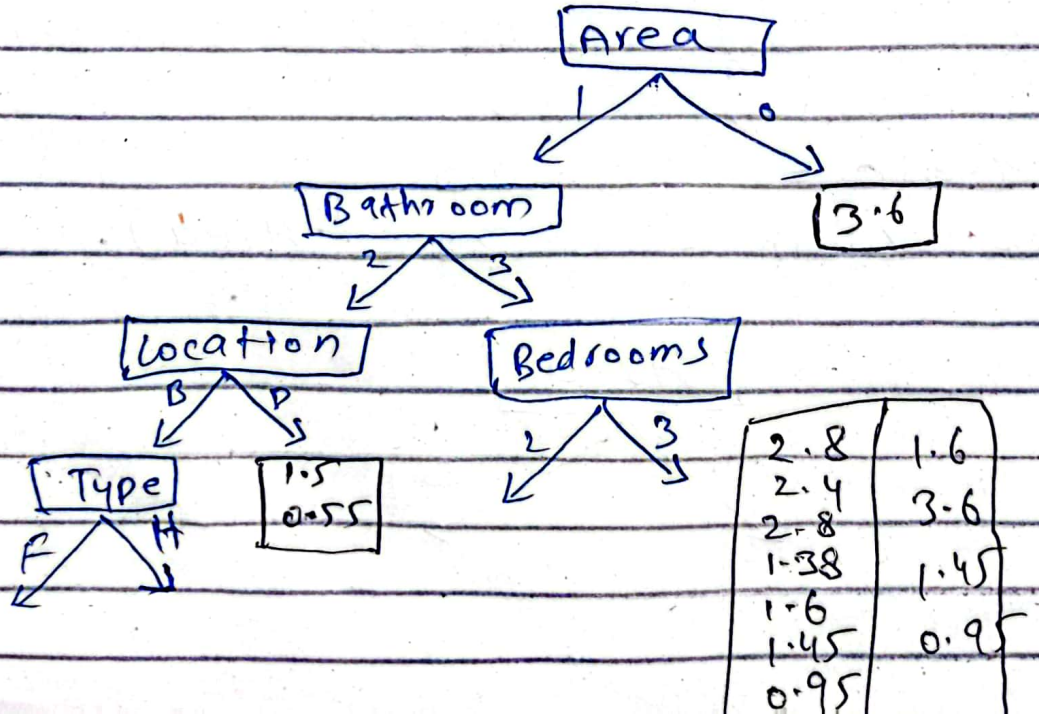
$\Rightarrow Var(S) = 0.6745$

(i) Bedrooms:-

$V(\text{Bedroom}-2) = 0$

$V(\text{Bedroom}-3) = 0.6745$

Gain = 0



2.8	1.6
2.4	3.6
2.8	
1.38	1.45
1.6	
1.45	0.95
0.95	

Date: Day: S M T W T F S S

⇒ XG BoosT Tree - I

→ Var(S) = 0.73558

(i) Bedrooms:

V(Bedrooms-2) = 0.3000

V(Bedrooms-3) = 0.6827

Gain = 0.1549

(ii) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0.6827

Gain = 0.1549

(iii) Type:

V(Flat) = 0.7400

V(House) = 0.7048

Gain = 0.0119

(iv) Location:

V(Bahria) = 0.6186

V(DHA) = 0.8445

Gain = 0.0146

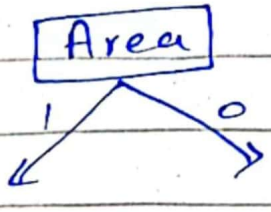
(v) Area:

V(≤236) = 0.5241

V(>236) = 0

Gain) = 0.2464

Max-Gain = 0.2464 (Area)



Date: _____ Day: M T W T F S S

Var(S) = 0.5241

(i) Bedrooms:

V(Bedrooms-2) = 0.3000

V(Bedrooms-3) = 0.4623

Gain = 0.1188

(ii) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0.4623

Gain = 0.1188

(iii) Type:

V(Flat) = 0.7400

V(House) = 0.0643

Gain = 0.0736

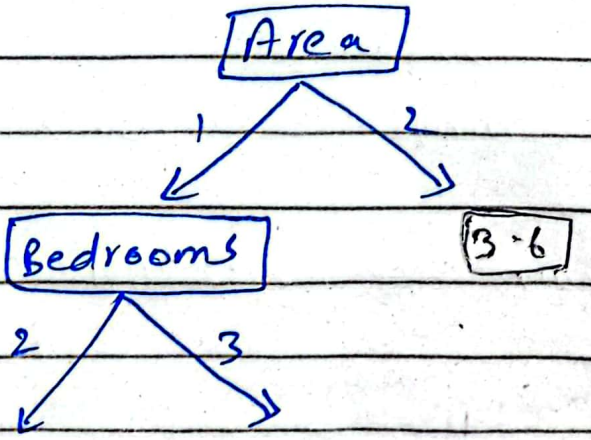
(iv) Location:-

V(Bahria) = 0.0814

V(DHA) = 0.8445

Gain = 0.0995

Max-Gain = 0.1188 (Bedrooms, Bathroom)



Date: _____ Day: M T W T F S S

→ Var (S) = 0.3000

(i) Bathroom:

V(Bathroom-2) = 0.3000

V(Bathroom-3) = 0

Gain = 0

(ii) Type:-

V(Flat) = 0.3000

V(House) = 0

Gain = 0

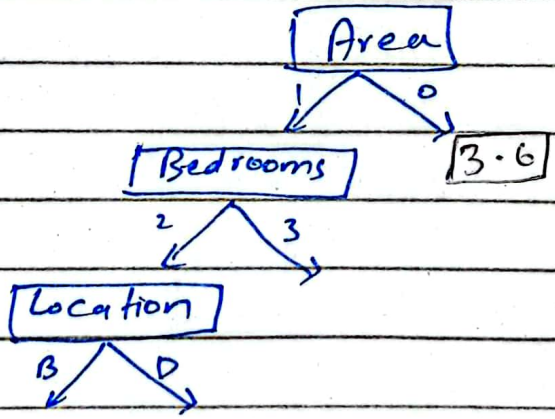
(iii) Location =

V(Bahria) = 0.2

V(DHA) = 0.2005

Gain = 0.1496

Max- Gain = 0.1496 (Location)



→ Var(S) = 0.4623

(i) Bathroom:

V(Bathroom-2) = 0

V(Bathroom-3) = 0.4623

Gain = 0

Date: _____ Day: M T W T F S S

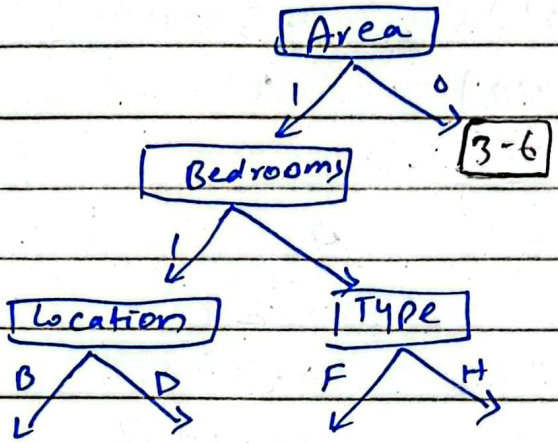
(ii) Type:

V(Flat) = 0.04

V(House) = 0.0643

Gain = 0.4076

Max-Gain = 0.4076 (Type)



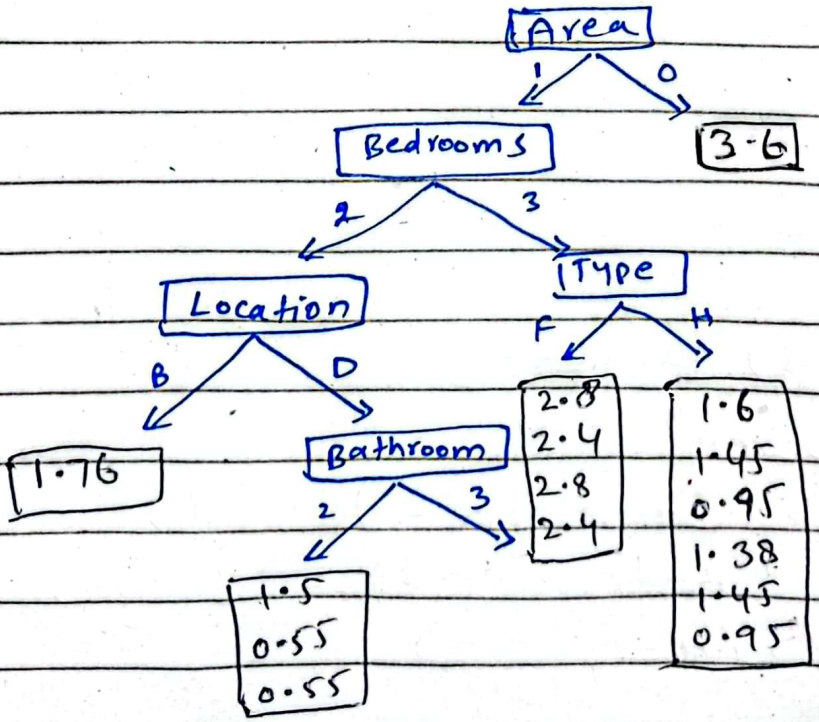
→ Var(S) = 0.08445

(i) Bathroom:

V(Bathroom-2) = 0.2005

V(Bathroom-3) = 0.04

Gain = 0.7357



⇒ XG BOOST Tree II:-

→ Var(S) = 0.78

(i) Type:- (Flat, House)

var (Flat) = 0.56

v (House) = 0.76

Gain = 0.012

(ii) Bedrooms:-

v (Bedrooms-2) = 0.3

v (Bedrooms-3) = 0.56

Gain = 0.18

(iii) Bathroom:-

v (Bathroom-2) = 0.3

v (Bathroom-3) = 0.56

Gain =

(iv) Location:-

v (Bahria) = 0.72

v (DHA) = 0.63

Gain = 0.03

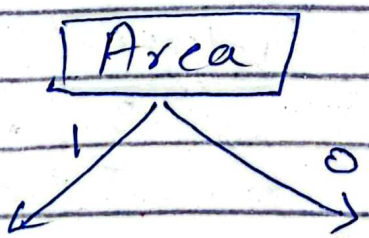
(v) Area:

v (≤236) = 0.53

v (>236) = 0

Gain = 0.28

Max- Gain = 0.28 (Area).



→ var(S) = 0.56

(i) Type:

v(Flat) = 0.76
v(House) = 0.06
Gain = 0.05

(ii) Bedrooms:-

v(Bedrooms-2) = 0.4
v(Bedrooms-3) = 0.31
Gain = 0.10

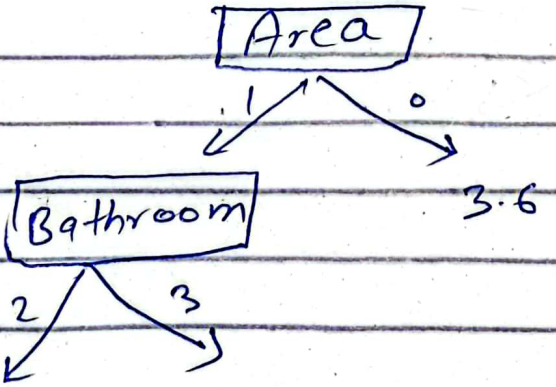
(iii) Bathroom:

v(Bathroom-2) = 0.4
v(Bathroom-3) = 0.31
Gain = 0.10

(iv) Location:

v(Bahria) = 0.109
v(DHA) = 0.64
Gain = 0.06

Max - Gain = 0.10 (Bedrooms, Bathroom)



→ Var(S) = 0.32

(i) Type:

v(Flat) = 0.0

v(House) = 0.0

Gain = 0.0

(ii) Bedrooms:

v(Bedrooms-2) = 0.0

v(Bedrooms-3) = 0.0

Gain = 0.0

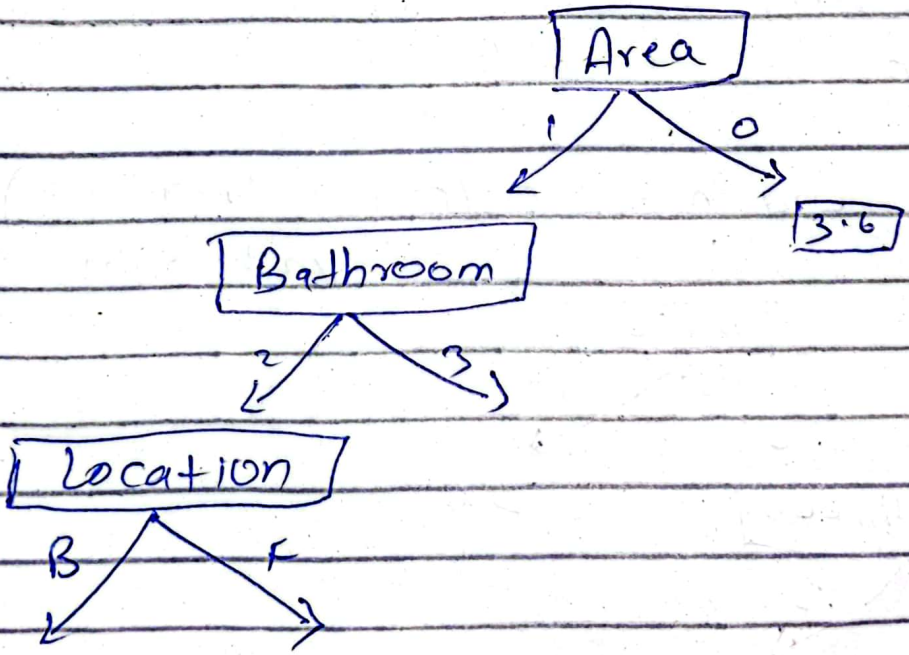
(iii) Location :-

v(Bahria) = 0.34

v(DHA) = 0.05

Gain = 0.04

Max - Gain = 0.04 (Location)



→ var(S) = 0.31

(i) Type:-

v(Flat) = 0.01

v(House) = 0.03

Gain = 0.29

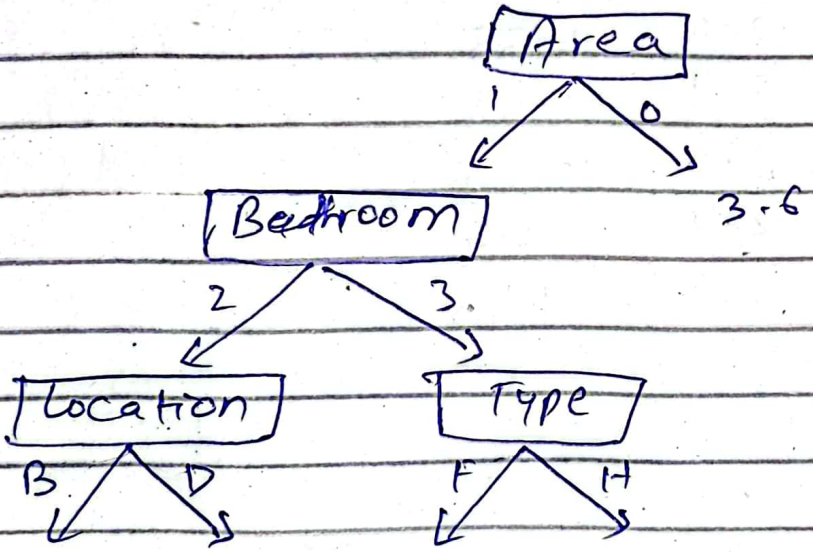
(ii) Bedroom:

v(Bedroom-2) = 0.0

v(Bedroom-3) = 0.0

Gain = 0.0

Max - Gain = 0.29 (Type).



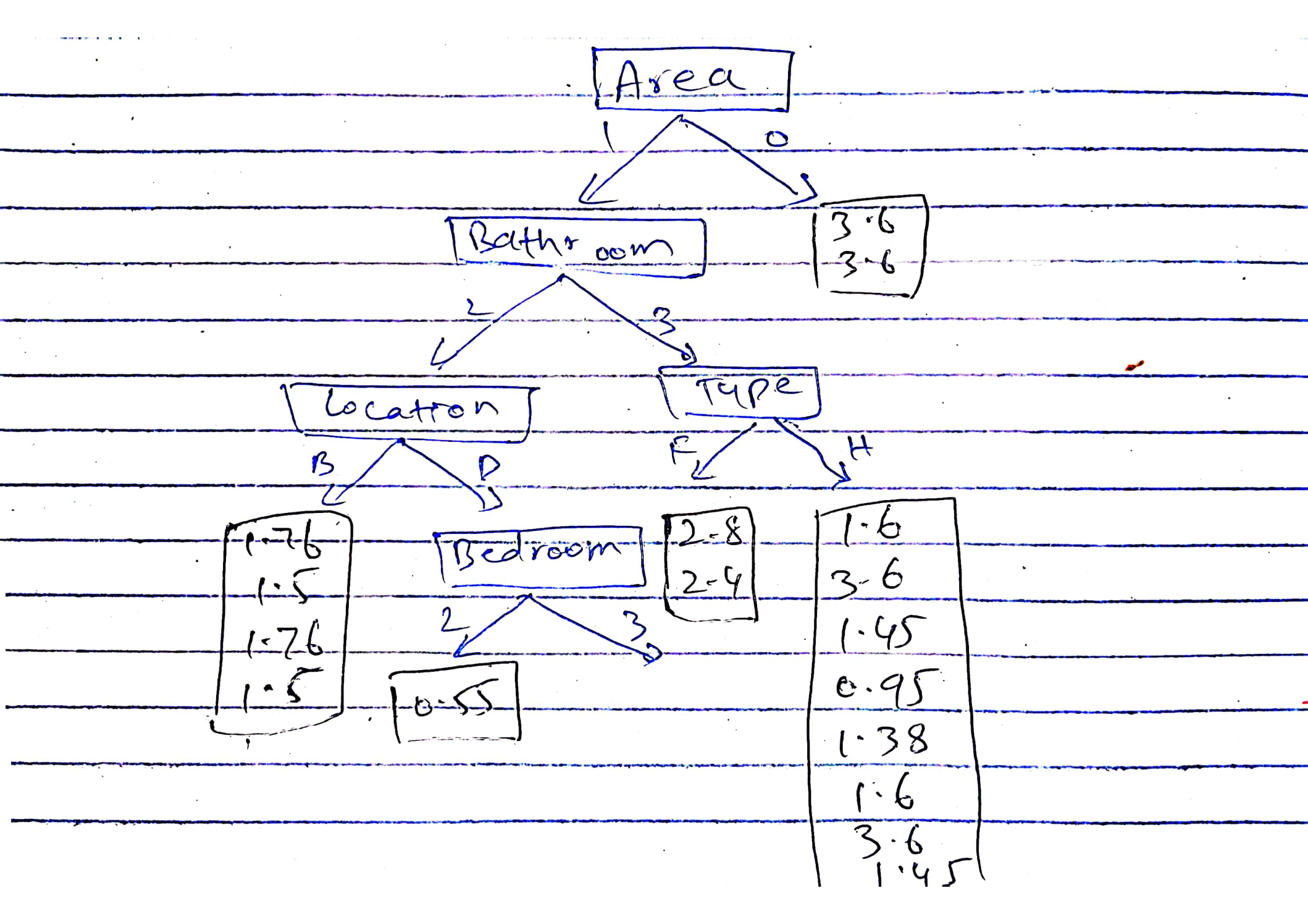
→ var(S) = 0.31

(i) Bedrooms:-

v(Bedrooms-2) = 0.45

v(Bedrooms-3) = 0.02

Gain = 0.55


Area
1
0
3·6
3·6
Bathroom
2
3
Location
Type
B
D
F
H
1·76
1·5
1·76
1·5
Bedroom
2·8
2·4
2
3
0·55
1·6
3·6
1·45
0·95
1·38
1·6
3·6
1·45

⇒ Mean squared Error (Regression Tree)

→ Training Data:

Actual Price	predicted Price
1.76	1.28
1.5	1.05
1.6	1.363
3.6	3.6
1.45	1.363
0.95	1.363
2.8	2.65
2.4	2.65
0.55	1.05
1.38	1.363
1.1	1.363
1.45	1.05
1.35	1.363
2.75	2.65
0.80	1.28

⇒ Test Data:

Actual price	predicted Price
0.77	1.28
0.70	1.28
0.81	1.28
1.8	1.28
3.35	2.65

$MSE_{Test} = 0.31556$

$MSE_{Train} = 0.09815$

As $MSE_{train} < MSE_{test}$ So, the model is High varience.

⇒ Mean Squared Error (Random Forest)

⇒ Training Data:- Tree - I

Actual price	predicted price
1.76	1.76
1.5	0.86
1.6	1.29
3.6	3.6
1.45	1.29
0.95	1.29
2.8	2.6
2.4	2.6
0.55	0.86
1.38	1.29
1.45	1.29
0.95	1.29
2.8	2.6
2.4	2.6
0.55	0.86

⇒ Test Data

Actual price	predicted price
0.77	1.29
0.76	1.29
0.81	1.29
1.8	1.29
3.35	2.6

$MSE_{Train} = 0.07656$

$MSE_{Test} = 0.33430$

As $MSE_{Train} < MSE_{Test}$ the model is High varience.

⇒ Mean Square Error (Random Forest Tree II)

→ Training Data

Actual Price	Predicted Price
1.76	1.8
1.5	1.02
1.6	1.8
3.6	3.6
1.45	1.8
0.95	1.8
2.8	1.8
2.4	1.8
0.55	1.02
1.38	1.8
1.6	1.8
3.6	1.8
1.45	1.8
0.95	1.8
2.8	1.8
	1.8

⇒ Test Data :-

Actual Data	Predicted Data
0.77	1.8
0.70	1.8
0.81	1.8
1.8	1.8
3.35	1.8

$MSE_{Train} = 0.53329$

$MSE_{Test} = 1.13070$

As $MSE_{Train} < MSE_{Test}$ then the model is High varience.

As $MSE_{train} < MSE_{test}$ So, the model is High varience.

⇒ Mean Squared Error (XG BOOST Tree - 1)

⇒ Training Data:-

Actual price	predicted price
1.76	1.76
1.5	0.86
1.6	1.29
3.6	3.6
1.45	1.29
0.95	1.29
2.8	2.6
2.4	2.6
0.55	0.86
1.38	1.29
1.45	1.29
0.95	1.29
2.8	2.6
2.4	2.6
0.55	0.86

⇒ Test Data

Actual price	predicted price
0.77	1.29
0.76	1.29
0.81	1.29
1.8	1.29
3.35	2.6

$MSE_{Train} = 0.07656$

$MSE_{Test} = 0.33430$

As $MSE_{Train} < MSE_{Test}$ the model is High varience.

⇒ Mean Squared Error (XG BOOST Tree - II)

⇒ Training Data

Actual price	Predicted Price
1.76	1.63
1.5	1.63
1.6	1.95
[redacted]	[redacted]
3.6	3.6
1.45	1.95
0.45	1.95
2.8	2.6
2.4	2.6
0.55	0.55
1.38	1.95
1.76	1.63
1.5	1.63
1.6	1.95
3.6	3.6
1.45	1.95

⇒ Test Data :-

Actual Price	predicted price
0.77	1.95
0.70	1.95
0.81	1.95
01.8	1.95
3.35	3.6

$MSE_{train} = 0.14783$

$MSE_{test} = 0.86790$

As $MSE_{Train} < MSE_{Test}$ then the Model is High varient.