1.) A cloth manufacturing company is interested to know about the segment or attributes causes high sale.

Approach - A decision tree can be built with target variable Sale (we will first convert it in categorical variable) & all other variable will be independent in the analysis.

Solution:-

**Business Problem:-** To build a decision tree for the Sale of a cloth manufacturing company

**Datasets:-**

Company.csv

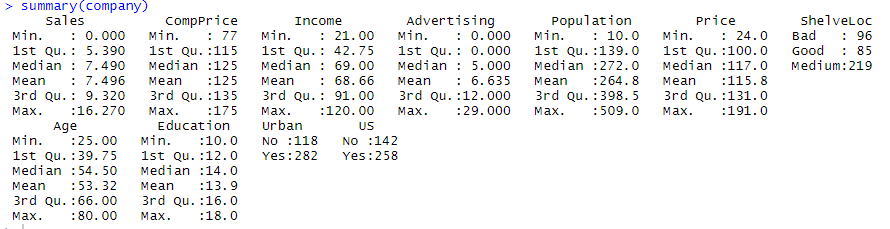
Dependent: CompPrice, Income, Advertising, Population, Price, ShelveLoc,

Age, Education, Urban, US

Independent: Sales

**EDA:-**

**Summary**



To convert sales into categorical variable, need to create bins. The values of sales variable is divided into three categories as “Low”, “Average”, “High” and attached as a new variable salesstatus to the company datasets

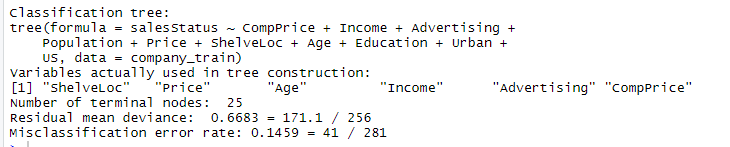
**Data Partitioning**

The datasets is divided into train and test data sets with 7:3 ratio.

**Model Building:**

The decision tree model is built on train dataset.

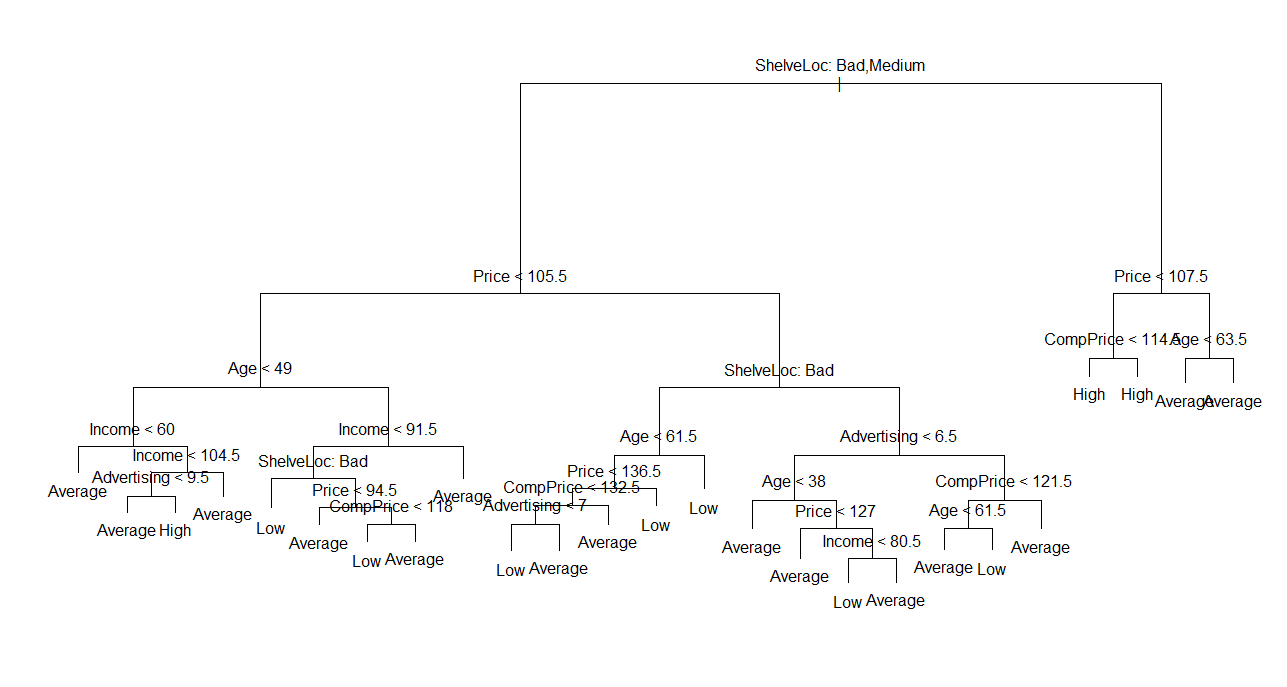
Model Summary



Error rate for the model is 14.59%

Accuracy = 100-14.59 = 85.41%

**Network Topology**



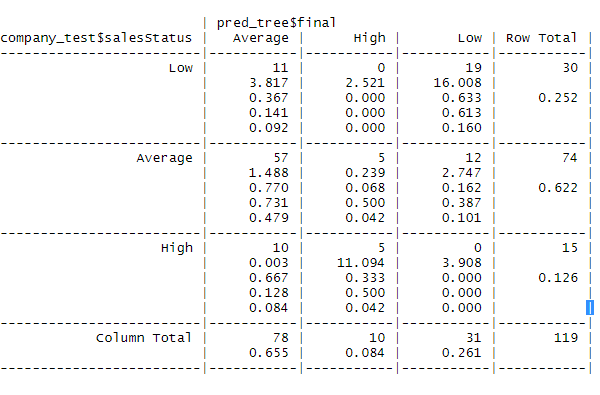
The above diagram represents the decision tree

**Evaluation:**

The model performance is checked on test data set.

Test Accuracy = 68.067%

**Cross tabulation**



Total data points correctly classified = 81

Total data points misclassified = 38

Error rate = 31.933%

In the above problem as the test accuracy is much less then train accuracy, we can say that it is facing the problem of overfitting.

2.) Use decision trees to prepare a model on fraud data treating those who have taxable\_income <= 30000 as "Risky" and others are "Good".

**Business Problem:-** To build a decision tree to identify risky or good.

**Datasets:-**

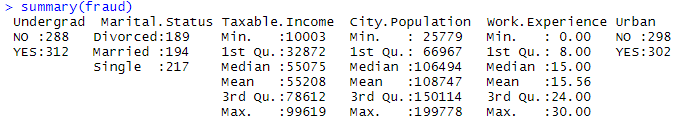
fraud.csv

Dependent: Undergrad, Marital.Status, City.Population, Work.Experience, Urban

Independent: Taxable.Income

**EDA:-**

**Summary**



Based on Taxable.Income, the data sets is divided as “Good” and “Risky

If Taxable. Income <= 30000 then “Risky” and Taxable.Income > 30000 then Good

A new variable “status” is added to the dataset containing values of Risky and Good.

**Data Partitioning**

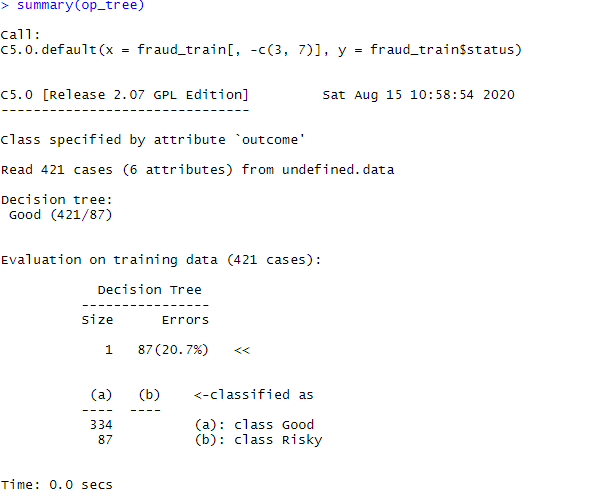
The datasets is divided into train and test data sets with 7:3 ratio.

**Model Building:**

The decision tree model is built on train dataset.

Model Summary





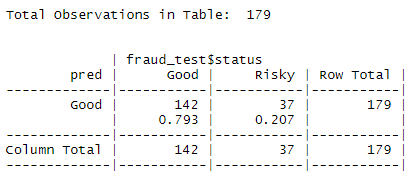


From the above plot 20 % of data is Risky and 80% is Good.

**Evaluation:**

The model performance is checked on test data set.

**Crosstabulation**

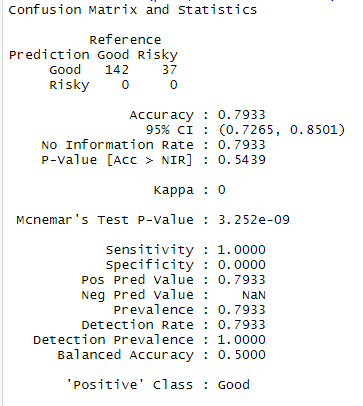


Total datapoints correctly classified = 142

Total datapoints misclassified = 37

Error rate = 20.67%

**Confusion Matrix**



**Accuracy = 79.53 %**