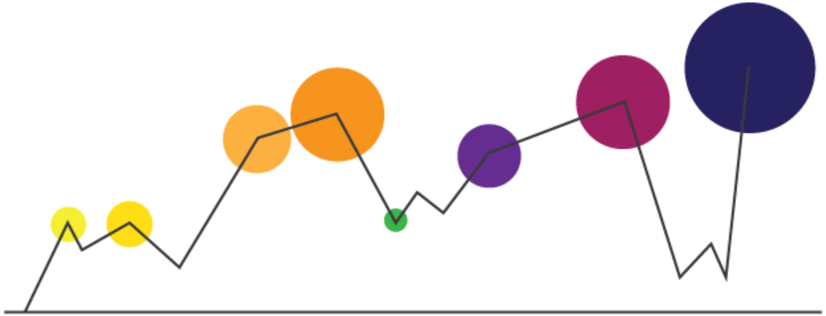


DECODING THE SCIENCE OF DECISION TREES! LEARN FROM EXPERTS

Agenda

Today we will take you through the following:

- ✓ The Classic Banking Challenge !!.. *Have you already guessed it??*
- ✓ The Available Options for Solution
- ✓ Why Decision Tree?
- ✓ How Decision Tree Methodology Works ?

The Classic Situation…

The Problem?

A bank wants to classify its future customers into two categories “Risky” and “Good” based on customer’s available attributes.

Let’s say a customer xyz has the following attributes. How will the bank know to which category this customer belong.

Undergrad	Marital Status	Taxable Income	City Population	Work Experience (Yrs)	Urban	Category
No	Married	98,727	1,01,894	14	NO	????

Let See Few More Cases..

A manager has to decide whether he should hire more human resources or not in order to optimize the work load balance

An individual has to make a decision such as whether or not to undertake a capital project, or must chose between two competing ventures

The Available Solution Options…

Algorithms that can help..

Such type of problems comes under **"classification"**

It is the separation or ordering of objects into classes

★ There are few techniques in classification method, like:

- ✓ Decision Tree
- ✓ Naïve Bayes
- ✓ k-Nearest Neighbor
- ✓ Support Vector Machine etc..

Why Decision Tree is Favorable..?

edureka!

Advantages of Decision Tree Methodology

	DT	NB	KNN	SVM
Simple visual representation of a decision situation	YES	NO	NO	NO
Easy to interpret and explain to executives (Non-programmers)!	YES	NO	NO	NO
Illustrates a variety of decisions and also the impact of each decision if different decisions were to be taken	YES	NO	NO	NO
Allow us to predict, explain, describe, or classify an outcome altogether	YES	NO	NO	NO
Help determine worst, best and expected values for different scenarios	YES	NO	NO	NO
Able to handle both numerical and categorical data	YES	NO	NO	NO

Decision Tree (DT)
Naïve Bayes (NB)
k-Nearest Neighbor (KNN)
Support Vector Machine (SVM)

Decision Tree Advantages..

Easy to interpret and explain to executives (Non-programmers)!

Decision Trees are

"white boxes" : The acquired knowledge can be expressed in a readable form,

while KNN,SVM,NB are

“black boxes”, :You cannot read the acquired knowledge in a comprehensible way

e.g. To Explain a suitable Weather Condition for Playing in Decision Tree format..

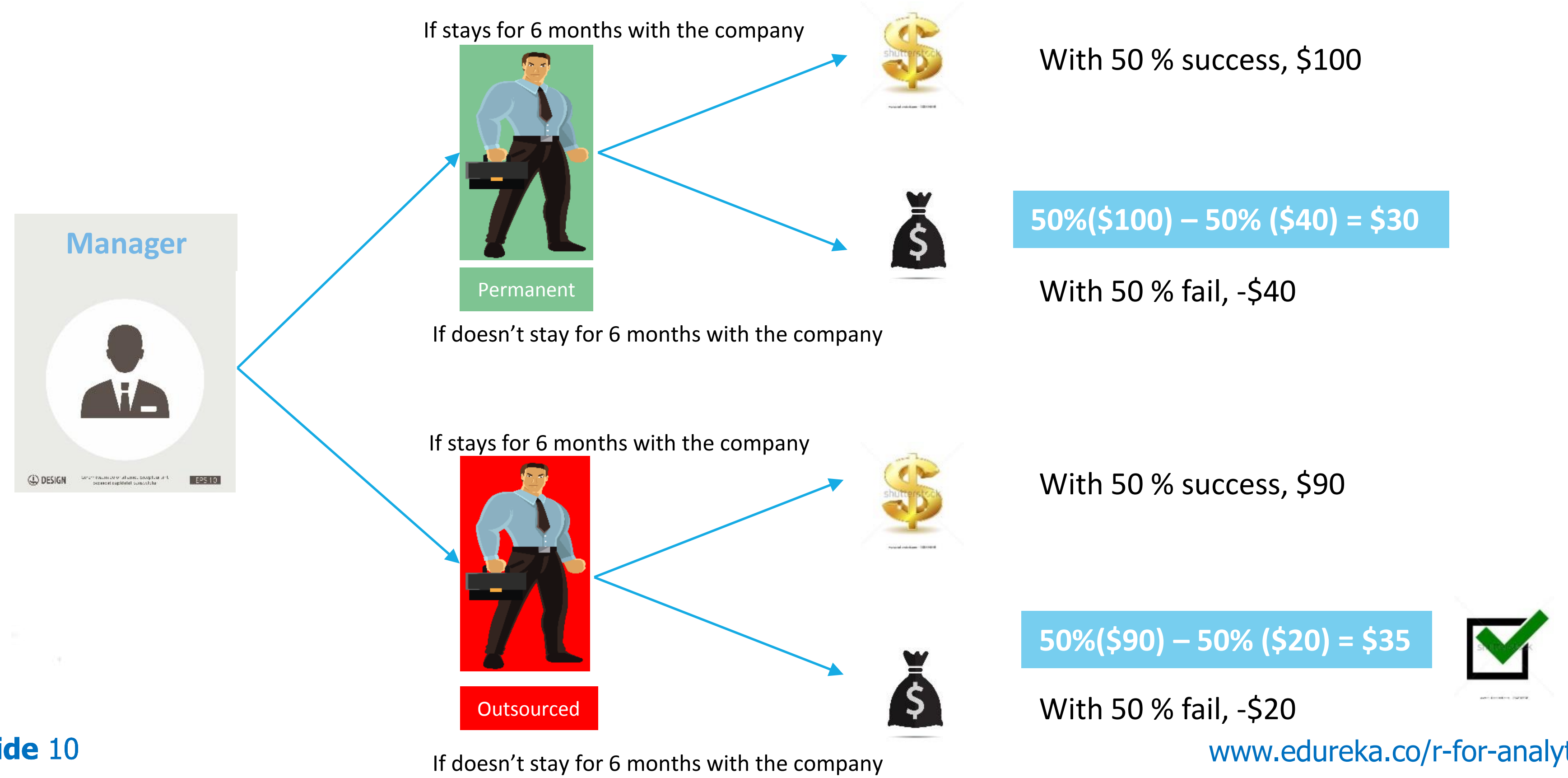
Cond. 1 **Cond. 2** **Cond. 3**

If weather is nice and wind is normal and the day is sunny then only **play** **(*Readable Format)**

Decision Tree Advantages..contd

edureka!

Illustrates a variety of decisions and also the impact of each decision if different decisions were to be taken



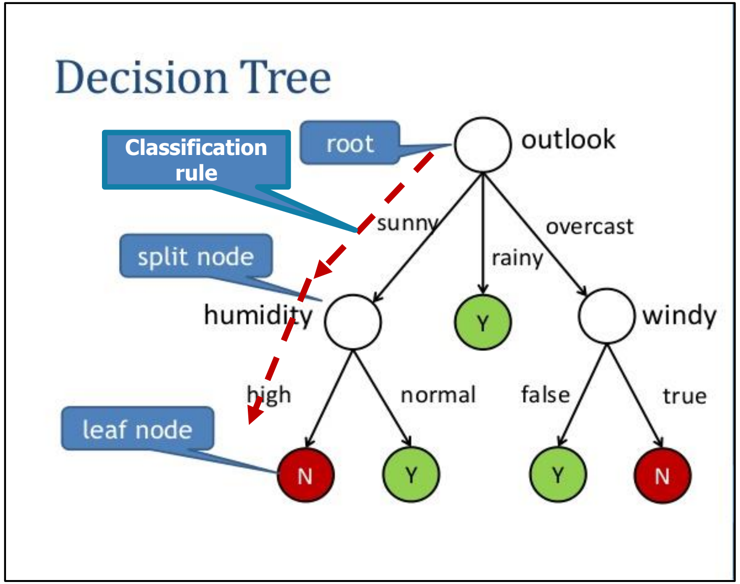
Let’s Understand it More

… What is Decision Tree?

Decision Tree

Decision Tree is a supervised rule based classification

- ✓ Flowchart - like tree structure
- ✓ The topmost node in a tree is the root node
- ✓ Each internal node denotes a test on an attribute, e.g. whether a coin flip comes up heads or tails
- ✓ Each branch represents an outcome of the test
- ✓ Each leaf node holds a class label (decision taken after computing all attributes)
- ✓ Paths from root to leaf represents classification rules



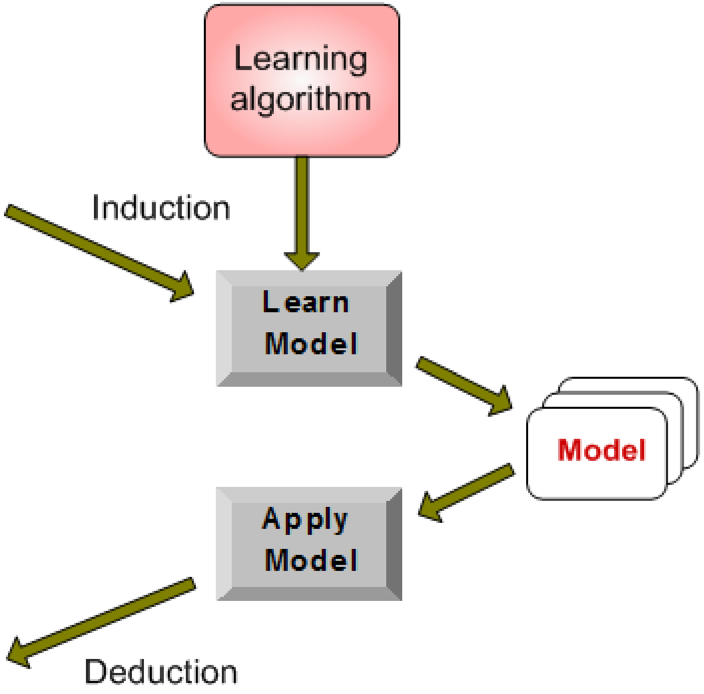
* * During tree construction, attribute selection measures are used to select the attribute which best partitions the tuples into distinct classes

DT Can Be used With Machine Learning

edureka!

When Coupled with Machine Learning, Decision Tree can be used for Prediction

Under -grad	Marital Status	Taxable Income	City Population	Work Experience	Urban	Category
Yes	Married	98,727	1,01,894	14	NO	Risky
No	Single	44,000	10,18,945	12	YES	Good
No	Divorced	50,000	10,15,845	14	YES	Good
No	Single	32,100	12,58,945	12	NO	Risky
Yes	Married	28,000	1,22,945	8	YES	Risky
No	Single	35,100	12,56,845	10	NO	Good
No	Divorced	38,100	18,95,945	7	NO	Good

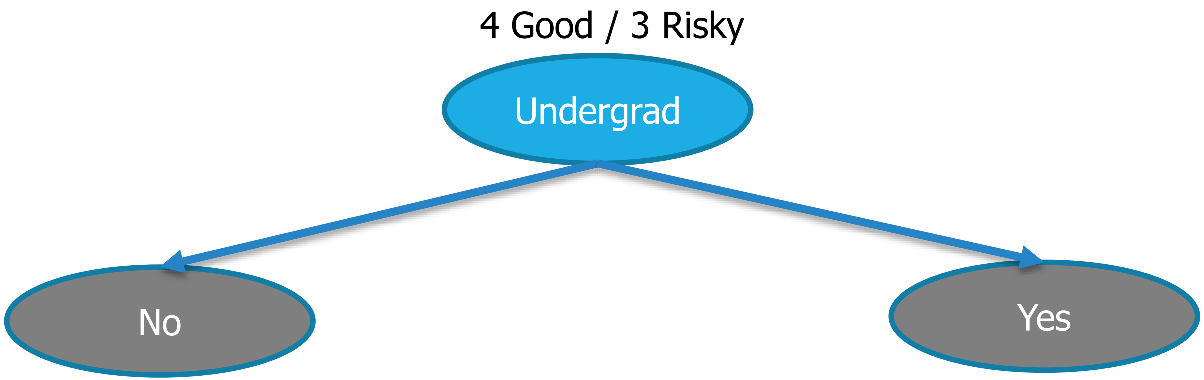


Under -grad	Marital Status	Taxable Income	City Population	Work Experience	Urban	Category
Yes	Divorced	98,727	1,01,894	14	NO	????

How it Works..

Let’s Build a Decision Tree Model !

Train Model (Build Tree)



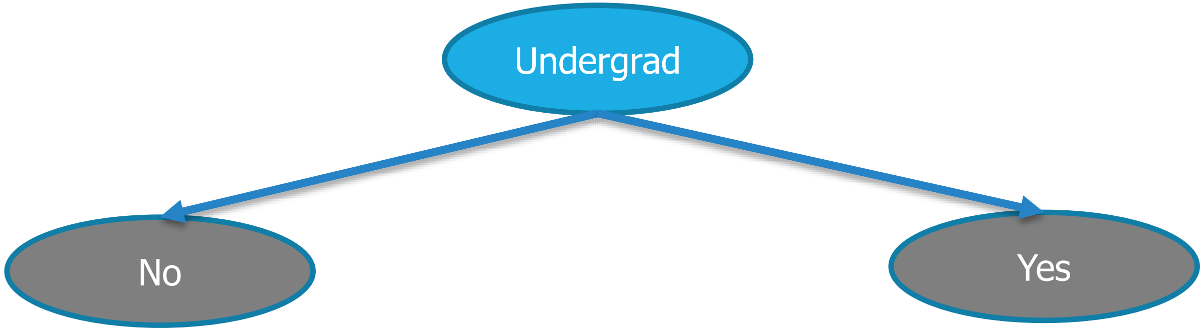
Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Single	44,000	1,01,8945	12	YES	Good
Divorced	50,000	1,01,5845	14	YES	Good
Single	32,100	1,25,8945	12	NO	Risky
Single	35,100	1,25,6845	10	NO	Good
Divorced	38,100	1,89,5945	7	NO	Good

Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Married	98,727	1,01,894	14	NO	Risky
Married	28,000	1,22,945	8	YES	Risky

Train Model (Build Tree)

edureka!

4 Good / 3 Risky



Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Single	44,000	1,01,8945	12	YES	Good
Divorced	50,000	1,01,5845	14	YES	Good
Single	32,100	1,25,8945	12	NO	Risky
Single	35,100	1,25,6845	10	NO	Good
Divorced	38,100	1,89,5945	7	NO	Good

4 Good/1 Risky

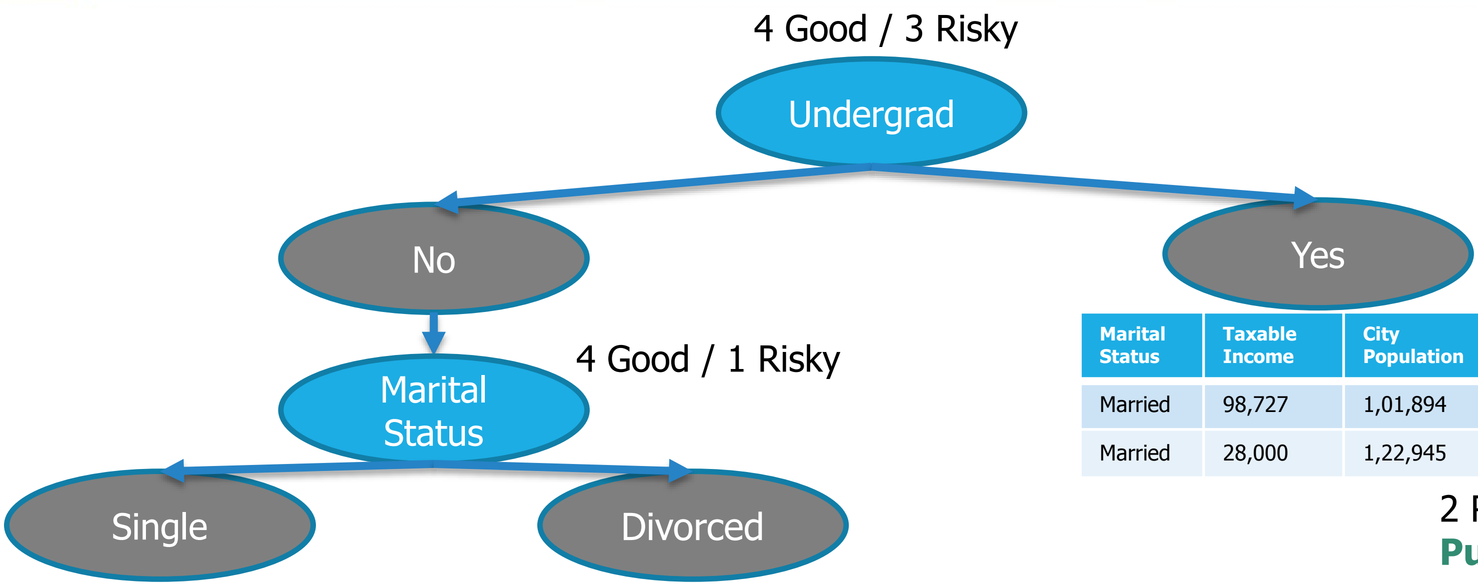
Split Further

Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Married	98,727	1,01,894	14	NO	Risky
Married	28,000	1,22,945	8	YES	Risky

2 Risky

Pure Subset

Train Model (Build Tree)



Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Married	98,727	1,01,894	14	NO	Risky
Married	28,000	1,22,945	8	YES	Risky

2 Risky

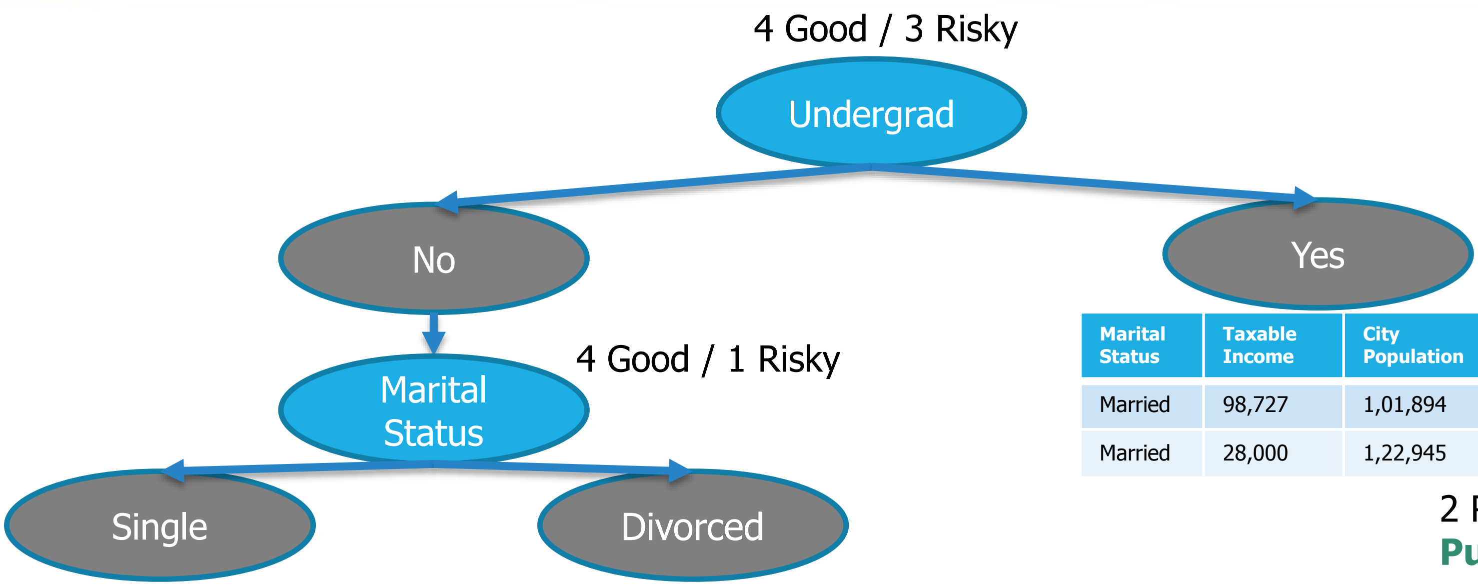
Pure Subset

Taxable Income	City Population	Work Experience	Urban	category
44,000	1,01,8945	12	YES	Good
32,100	1,25,8945	12	NO	Risky
35,100	1,25,6845	10	NO	Good

Taxable Income	City Population	Work Experience	Urban	category
50,000	1,01,5845	14	YES	Good
38,100	1,89,5945	7	NO	Good

Train Model (Build Tree)

edureka!



Marital Status	Taxable Income	City Population	Work Experience	Urban	category
Married	98,727	1,01,894	14	NO	Risky
Married	28,000	1,22,945	8	YES	Risky

2 Risky
Pure Subset

Taxable Income	City Population	Work Experience	Urban	category
44,000	1,01,8945	12	YES	Good
32,100	1,25,8945	12	NO	Risky
35,100	1,25,6845	10	NO	Good

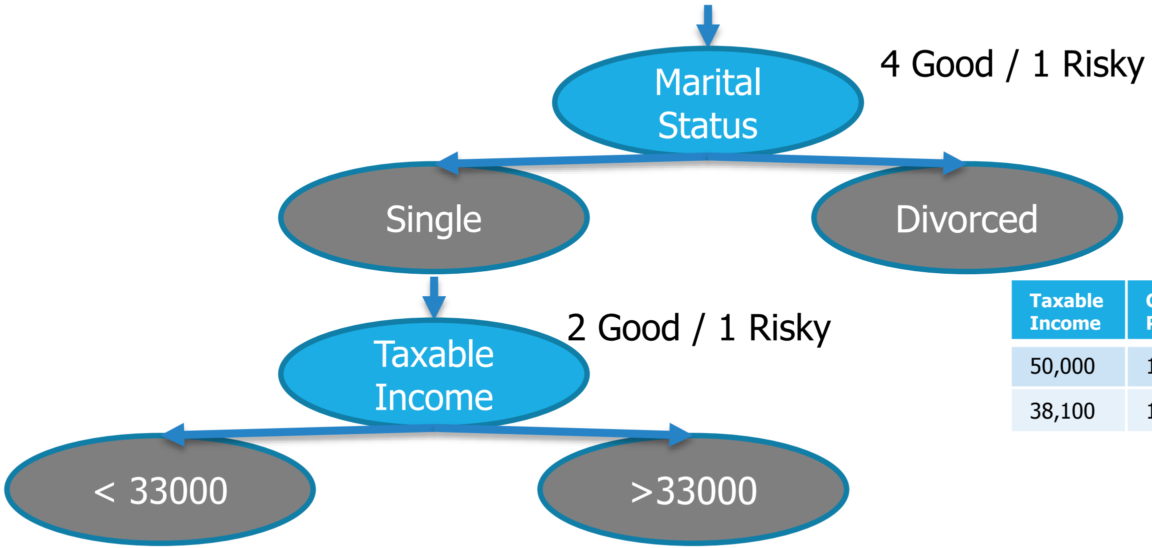
2 Good/1 Risky
Split Further

Taxable Income	City Population	Work Experience	Urban	category
50,000	1,01,5845	14	YES	Good
38,100	1,89,5945	7	NO	Good

2 Good
Pure Subset

Train Model (Build Tree)

edureka!



Taxable Income	City Population	Work Experience	Urban	category
50,000	1,01,5845	14	YES	Good
38,100	1,89,5945	7	NO	Good

2 Good

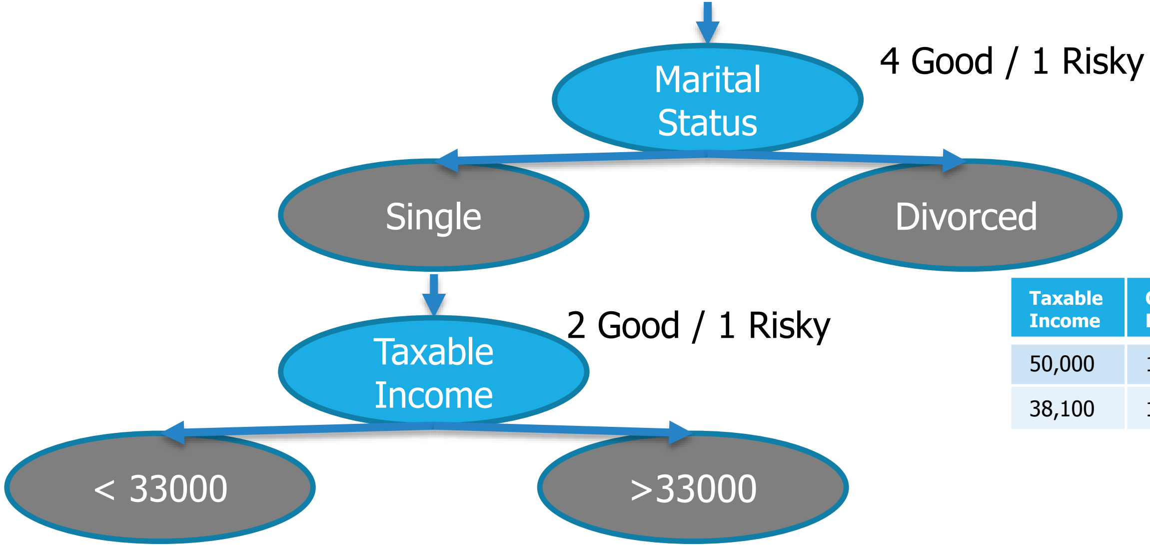
Pure Subset

Taxable Income	City Population	Work Experience	Urban	category
32,100	1,25,8945	12	NO	Risky

Taxable Income	City Population	Work Experience	Urban	category
44,000	1,01,8945	12	YES	Good
35,100	1,25,6845	10	NO	Good

Train Model (Build Tree)

edureka!



Taxable Income	City Population	Work Experience	Urban	category
50,000	1,01,5845	14	YES	Good
38,100	1,89,5945	7	NO	Good

2 Good
Pure Subset

Taxable Income	City Population	Work Experience	Urban	category
32,100	1,25,8945	12	NO	Risky

1 Risky
Pure Subset

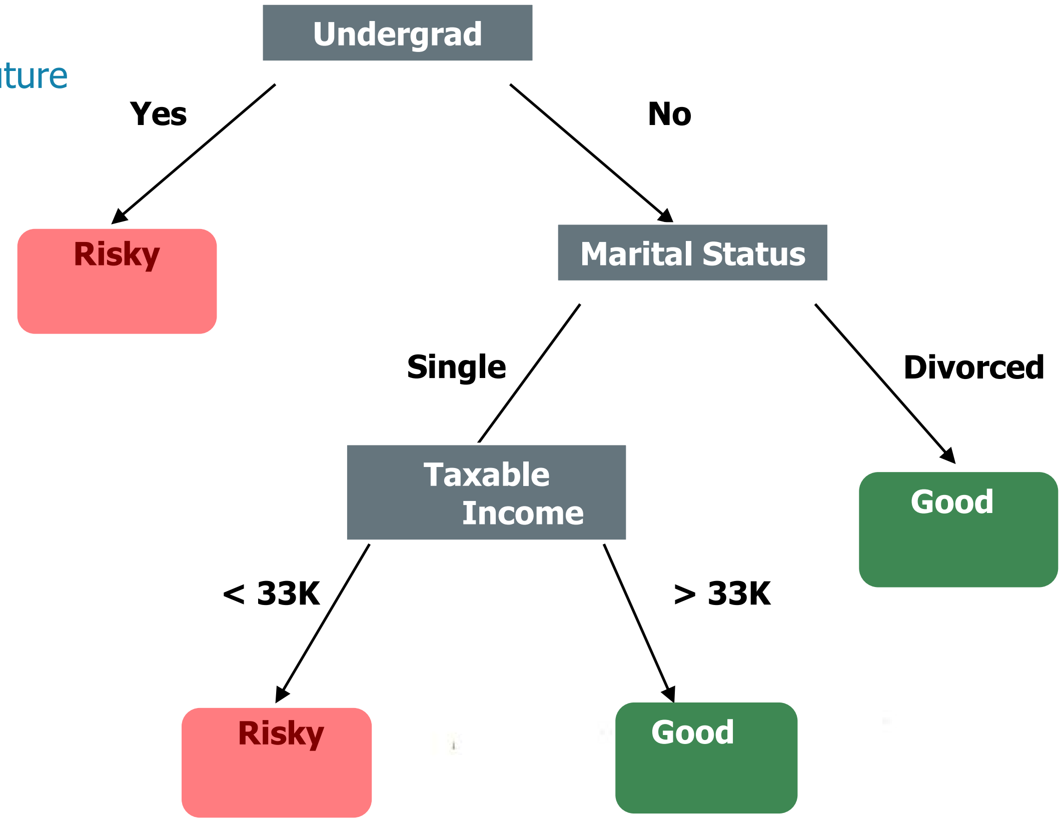
Taxable Income	City Population	Work Experience	Urban	category
44,000	1,01,8945	12	YES	Good
35,100	1,25,6845	10	NO	Good

2 Good
Pure Subset

The Final Built Model

Here is a trained model that will help us in future

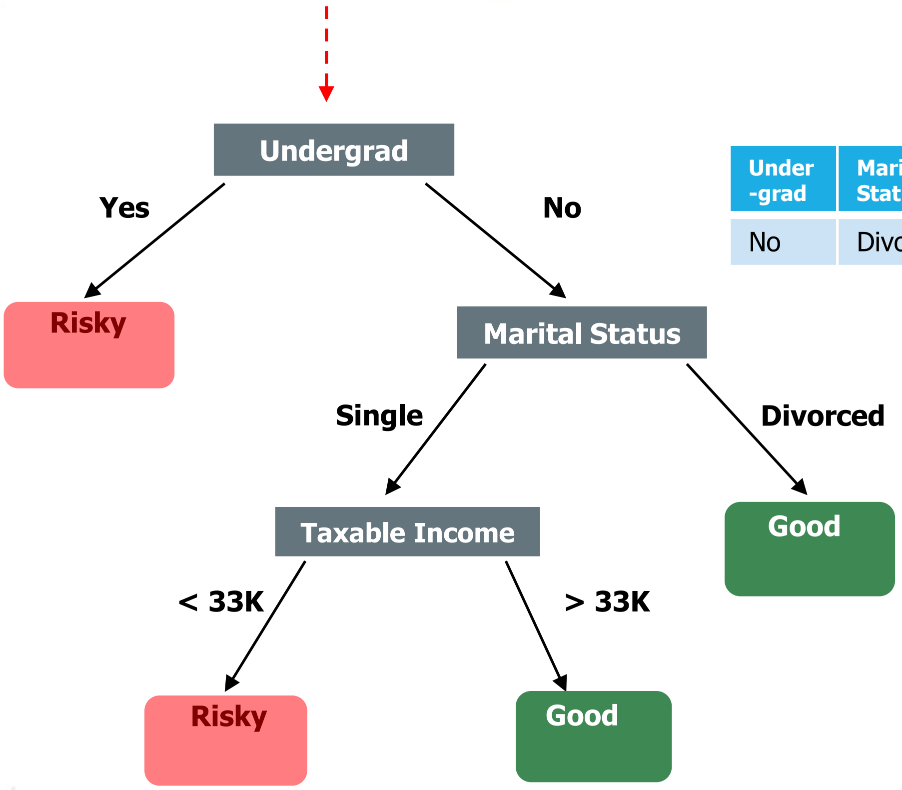
"**Classification**"



Let's Use our Model..

Start from the root of tree

edureka!



Under -grad	Marital Status	Taxable Income	City Population	Work Experience	Urba n	category
No	Divorced	98727	101894	14	NO	????

Test Data

How It Works

Under-grad	Marital Status	Taxable Income	City Population	Work Experience	Urban	category
No	Divorced	98727	101894	14	NO	????

Undergrad

Test Data

How It Works

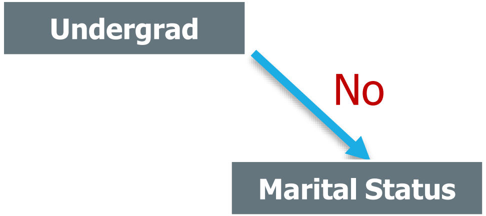
Under-grad	Marital Status	Taxable Income	City Population	Work Experience	Urban	category
No	Divorced	98727	101894	14	NO	????

Undergrad

No

Test Data

How It Works



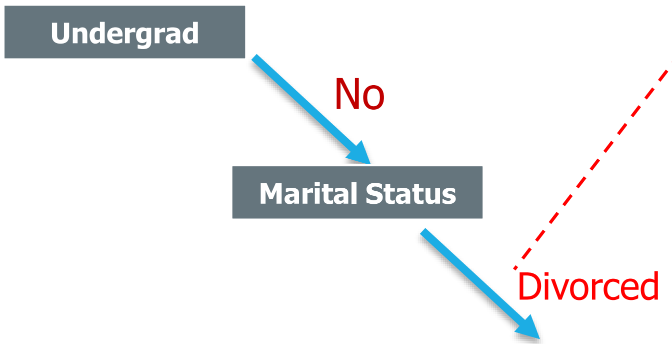
Marital Status	Taxable Income	City Population	Work Experience	Urba n	category
Divorced	98727	101894	14	NO	????

Test Data

How It Works

edureka!

Marital Status	Taxable Income	City Population	Work Experience	Urba n	category
Divorced	98727	101894	14	NO	????

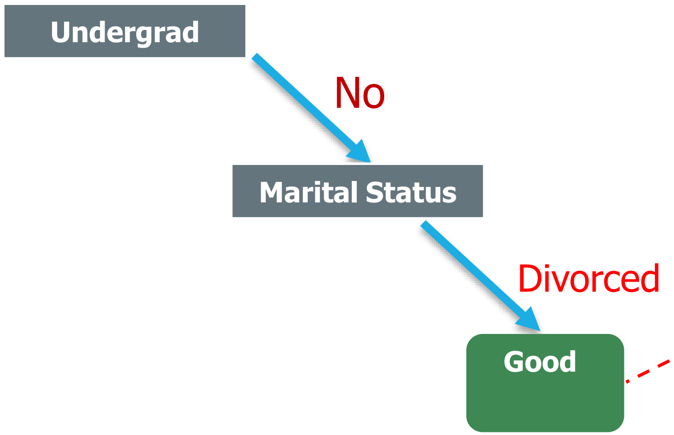


Test Data

How It Works

Marital Status	Taxable Income	City Population	Work Experience	Urba n	category
Divorced	98727	101894	14	NO	Good

Test Data



The customer lies in the "Good" Group and he can be considered for all policies for a good customer like more loans can be provided to him, his credit card limit can be increased etc..

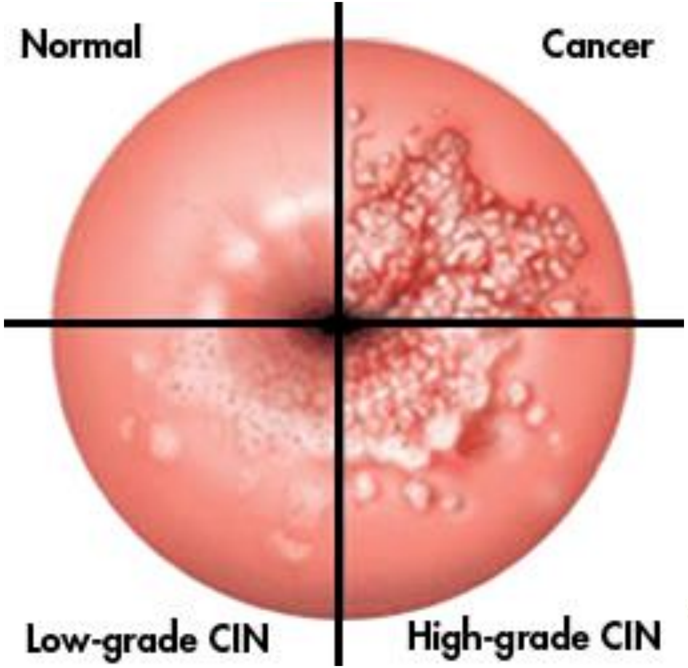
Demo

Real Life Application

Application of Decision Trees

edureka!

Predicting tumor cells as benign or malignant!



Banks Using For Classifying credit card transactions !

edureka!

Categorizing news stories as finance, weather etc…

QUESTIONS

Conclusion!

edureka!

When to Apply Decision Tree ??

→Whenever you are making a future complex decision

→When you have are just experimenting with the decisions and you want to evaluate and visualize your decision and the impact

→When you want to present your decision and its comparison with other decisions on the same problem

Survey

Your feedback is vital for us, be it a compliment, a suggestion or a complaint. It helps us to make your experience better!

Please spare few minutes to take the survey after the webinar.

edureka!