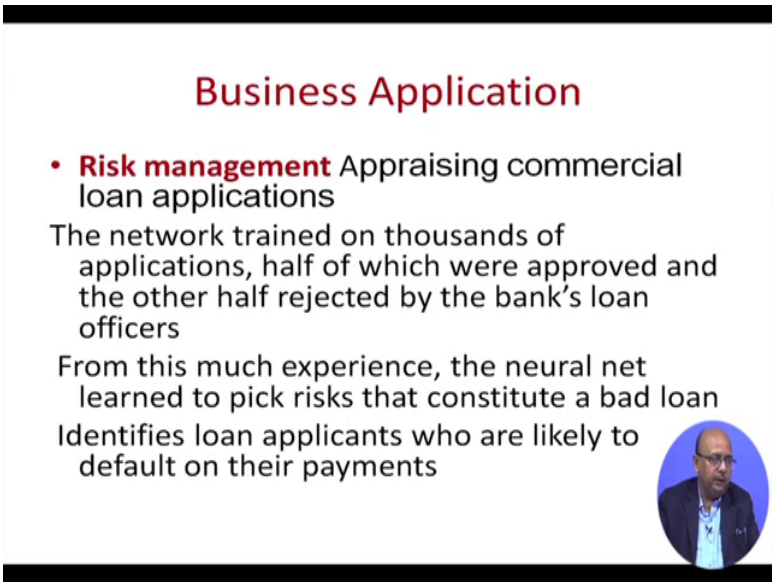
Knowledge Management
Prof. K B L Srivastava
Department of Humanities and Social Science
Indian Institute of Technology - Kharagpur

Lecture – 27
Inferences from Data (Contd.)

Okay, so now we are going to look into business applications of Neural Network and other things. Now when it comes to business applications we have taken certain examples in different business contexts.

(Refer Slide Time: 00:32)



And then we will see that how it can be done one example that is taken here is that how you go about risk management okay, like when a bank is going to grant you loan and for different purposes so you know that when the bank is granting loan many times you find that people are defaulting and not paying their loan that all are there times and they are becoming defaulters.

So it becomes a liability for the bank so suppose they want to mitigate the effect or risk and see that there are reduced risk when they are going to dispose loan applications so if you are going to use this network neural network then you have to see that in what way can proceed so one example that I have taken here is risk examples, where you are going to evaluate commercial loan applications in this case suppose now you find that okay.

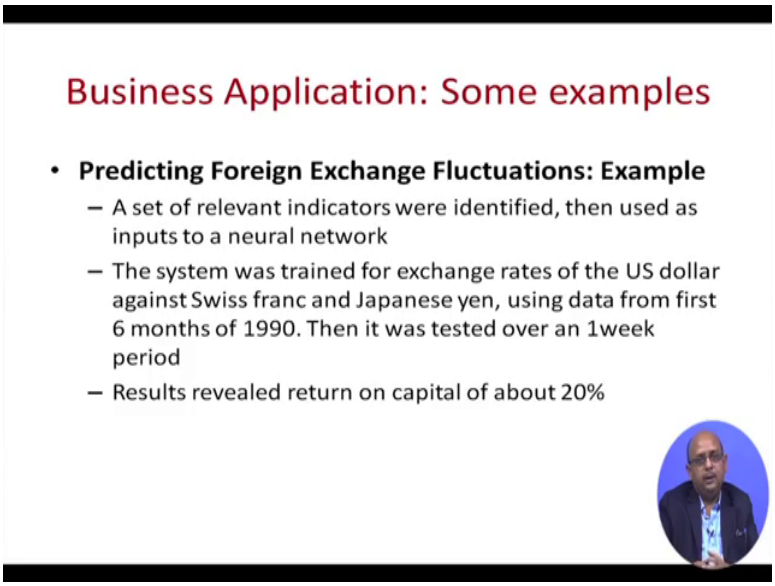
Number of people who have applied for loan some of them were approved as some of them

were not approved right, so you it means that there are two categories that you have created now loan that have been approved and loans that have not been approved okay. And this is based on what the judgment and the experience of certain procedures or rules that are followed by the bank employees or the bank people in approving or rejecting the loan applications.

So from there once you have bought this then you also see that those have been accepted for granting loan how many of them are going to default at later stage okay. Now which is going to create a risk because if you look at this a loan default means it is a bad loan right. So when it is approved and you are not able to pay since they have met these guesses or estimation based on certain criteria and you have been granted loan.

So but still there is a risk because you have defaulted on paying the loan right. So you can find out that what needs to be done okay.

(Refer Slide Time: 02:24)



So you are going to use these neural network see that what is the percentage of defaulters even after approval of the loan okay. Then you can further go that okay what are the factors which are responsible? What is their weight age okay? And then what kind of decisions you are going to take based on that when you are going to take the next round of loan applications. So these are basically business decisions which can be used.

Like suppose you know that changes in the political economic situation and the application that is there when you want to predict that how much fluctuations are going to be there in the

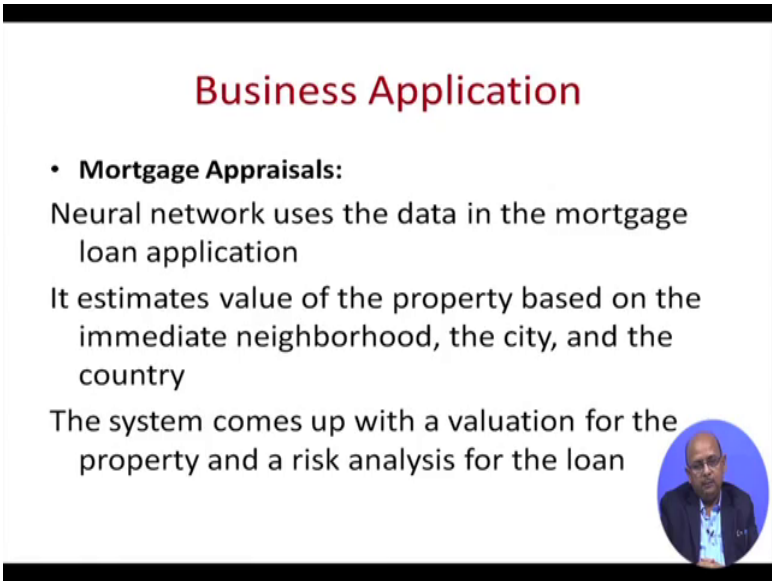
foreign exchange. Now the Foreign Exchange fluctuation is based upon the economy of a particular Country, for examples if you look at the rupees and if you convert into dollar you know that rupee value keep on changing okay.

Sometimes it is sixty eight rupees sometimes it is fifty, if sixty five rupees. So there is a fluctuation in the markets so far as Foreign Exchange is concerned. So what you are going to do, you are going to see that what are the factors or inputs that you could be used as a predictor of fluctuations so they are going to act as a input for you right. Then you are going to develop a system to see that how the exchange rate works and then you are going to put this data right.

And see that what is happening and that how much fluctuation is happening okay and then you can see that okay because of this foreign and foreign exchange fluctuations how much you are earning or how much you are losing right. So they are related to the applications of these neural networks in different cases another example that I have taken here is a Mortgage Applications right you can also use these data in Mortgage Applications, for example someone is going to buy home.

So what you do in the name of mortgage you take the history right unless they pay the entire loan you are not going to return the registration paper and these kind of things right.

(Refer Slide Time: 04:21)

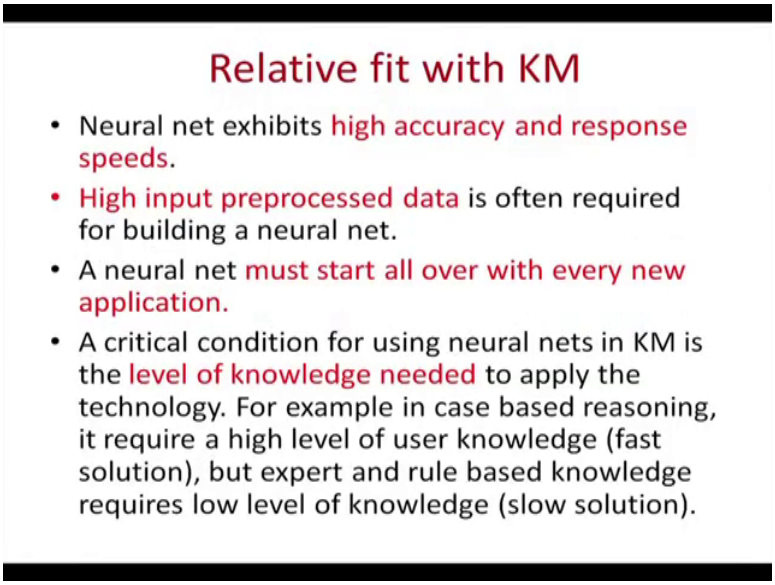


And so before mortgaging what you do you are going to take certain decisions based on what okay. Number of factors okay maybe the city, what is the property of the value, what is the

market value actually, what is a circle rates, okay what are the immediate number neighborhood because all these factors are going to decide about the price of the value right.

So based on this variables or inputs you can develop a system and then you are going to enquire certain thing from the data and you can also calculate the risk that is associated with giving loans right. So these are there are different application that can be taken up.

(Refer Slide Time: 04:58)



Now another issue is that how it fits with the knowledge management system. Since it is more objective data that you get and you are going to use sophisticated statistical tools for the purpose okay. These neural networks basically are related with high accuracy and responsibilities because if the data is correct then you get better results, correct results and the response speed is response speed means the date the results come very fast.

Because you have already programmed it into software and then you just put the data into the software and you get the results. Right if the data is accurate then you get correct a results out of it right. So you have input pre process data that is very, very important for example you need data on all inputs which you are going to put it in the equation are those variables which are to be process for the output right.

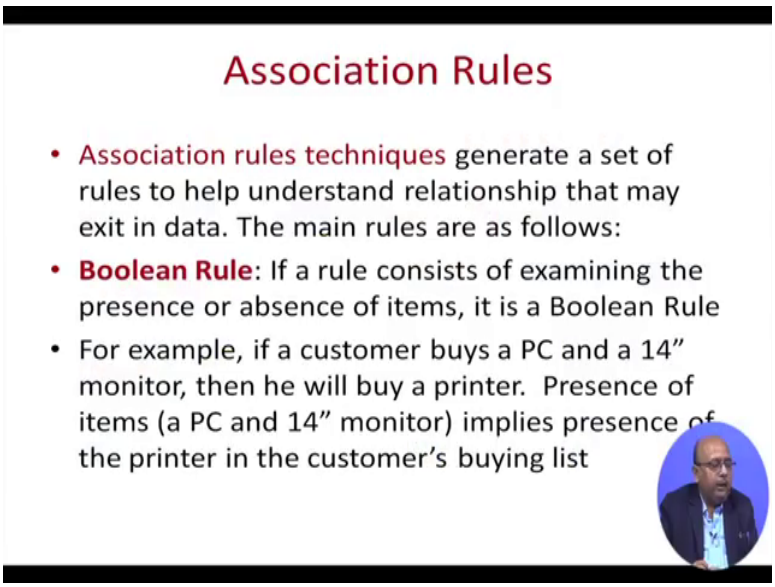
So that is the first stage and then you have to see that all of them are going to be put in the equation at the same time because you are not looking into one particular factor but all the factors all together so you must start all over with every new application for each application you are going to run the program okay, to see whether the loan application is to be rejected or

accept it right.

Now but the most critical factor is that what is the level of knowledge that is associated with the input. It is not the technology which is important but it is the level of knowledge which is associated with each input right which is going to give you the desired result right for example if you go for case based deciding you use high level of knowledge so you get fast result but if you are using expert or rule-based knowledge solutions are slow right.

So you have to see that the input variables in a neural network are very, very correct the high level of knowledge associated so that when that is put into the program or statistical rules it gives you better results.

(Refer Slide Time: 07:03)



We also discussed some rules here like Association Rules how are you going to associate say these inputs with the output like when you are going to mortgage loan applications you are going to consider certain input variables base on which you are going to decide the property of the value of the property right.

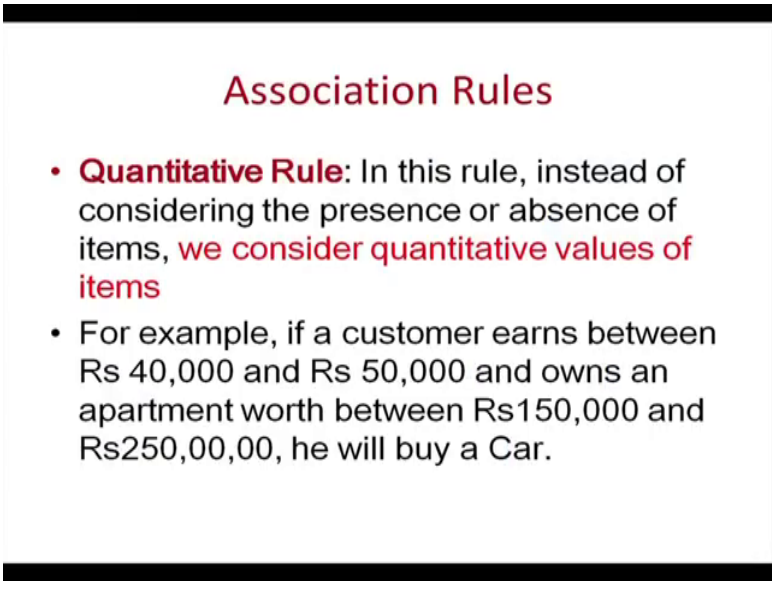
So you are going to use a set of rules to understand the relationship which is going to exist in the data right and it is very, very important for you to understand these rules like most common rules is Boolean rules if rule consists of examining the presence or absence of items, it is a Boolean Rule whether it is required or not for example if you identified six input variables whether all the six input variables are important or only some of them are important.

So you need to identify those critical variables which are going to predict okay and you can even leave out those which are insignificant or those not critically relate to the output right, one example that is given here is a very simple example that if a customer buys a PC and a fourteen inch mini monitor, then he is going to buy a printer right. So what is the input that PC and fourteen inch and based on what you are going to predict buying a printer that is the behavior right.

So you have made a rule that this is happening then this will happen right a person has a PC and a computer and monitor then there is a likely that he is going to have a monitor sorry a printer now you can add other rules also for example if he has PC with a monitor and he also has Internet connection right, then he is likely to buy a printer so you can identify these rules and if you think that if these rules are available and then you are going to see whether these items are there and this is going to be considered.

And these rules are not going to follow then it is not going to be considered right.

(Refer Slide Time: 08:59)

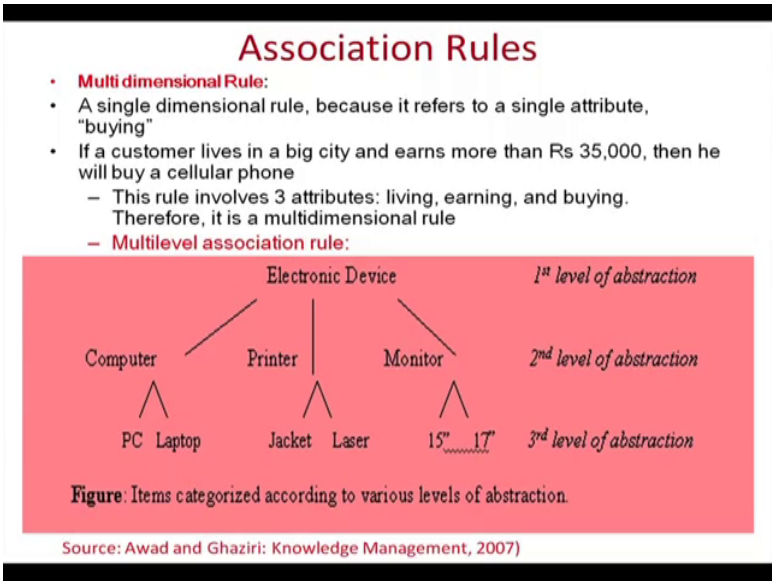


Then Quantitative rules right we basically look into the quantitative value of the item for example if you think the customer earns between forty thousand and fifty thousand and owns a apartment worth between say fifteen lakh to twenty five lakh then he is going to buy a car okay. So this is a more quantitative approach compared to what we have talked about earlier that is associational if you have a PC you need a printer right.

So you are going to associate a PC with the printer now hear what you are going to do it is

not simply a absence or presence of a item but you are going to see that how they all can be related so you are saying you are earning is this much then you are going or you are likely to buy a car. So as a seller I am going to approach you because if you fall in this category so then you are going to be our potential buyer right.

(Refer Slide Time: 09:55)



Now there could be single dimensional rule because you are going to look into one attribute or you can also have more attributes at different levels right for example you are living the a big city and more than thirty five thousand then you are going to buy a phone right. So this rule involves three attributes living that is you are living in a big city and if you are not living in a small city they will not going to buy it.

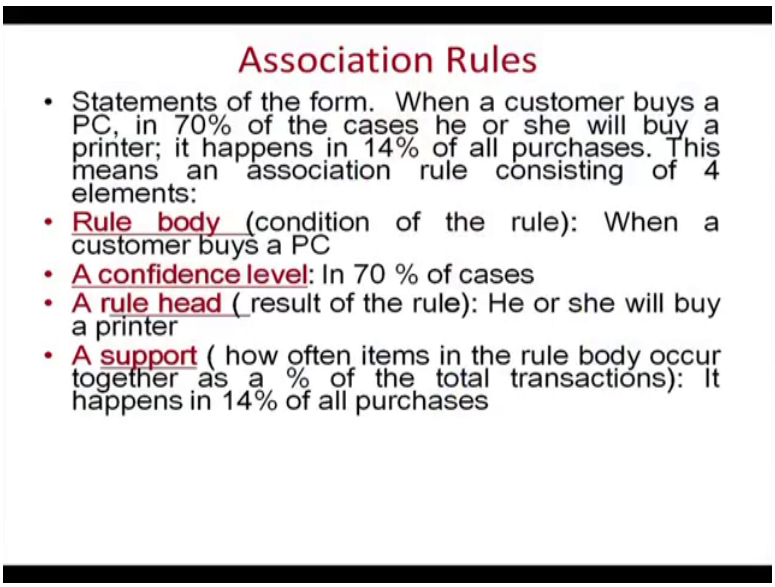
Then earning if you earning is more than thirty five thousand then you are going to buy it right and then you are going to do it so this is a Multi dimensional Rule. Now in a Multi dimensional Rule you are going to have multi level association to different level of associate now look at this example you are looking for electoral electronic devices that is the first level of abstraction, second level of abstraction could be PC, Printers and Monitor right.

Now in computers there are two options that you can have and then you move to another level of abstraction that whether you are going to buy a personal computer or whether you are going to buy a laptop. Now printer again you are going to have two options that you are going to buy inject printers or whether you are going to buy laser printers right. Then monitor whether you are going to buy fifteen inch or below fifteen inch or seventeen inch and above right.

So these things are categorized at a different level of abstraction so it becomes easy for you to decide say for example he needs a electronic device he is going for a computer that could be a PC and he is looking for a Laser printer and because he has a seventeen inch monitor right, you can further go for it. For example if you have Internet connection okay whether he have internet connection or he do not have an internet connection.

So that kind of abstraction has to be put into either the first level or the second level right. So the then you go at the fourth level of abstraction also right. So based on that you are going to associate whether the person is going to buy certain things or not when you are going to take a decision you make certain rules and before classifying your classified within that rule then you are going to do it otherwise not so the association rules right.

(Refer Slide Time: 12:01)

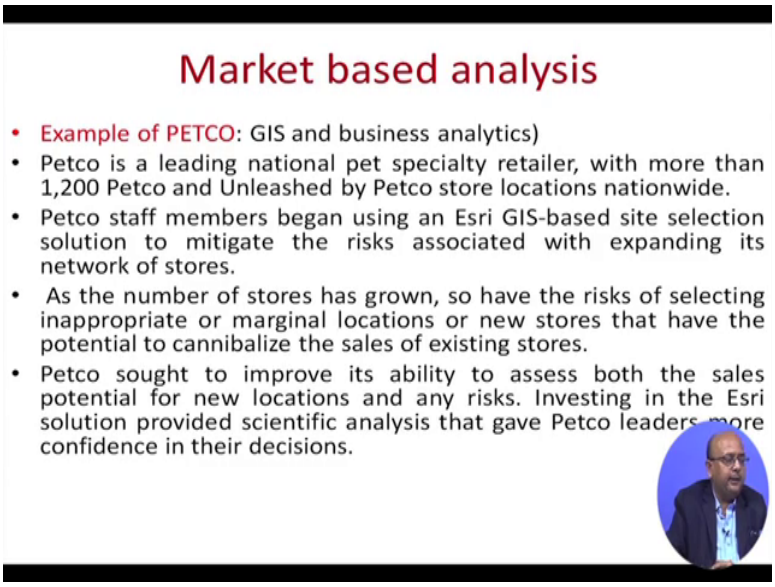


Look at this the four things that need to be considered it has four elements right. When a customer buys the PC in seventy percent of the cases he or she will buy a printer in the happens in all fourteen percent of all purchases. This means that an association rule consist of four element these are the four elements that we need to consider first is rule body that is when you are going to buy a PC that is must this is the first condition of the rule.

Then you have a confidence level and that is that in seventy percent of the cases it is likely that they are going to buy it right in thirty percent cases it is not going to happen, then result of rule that whether he is going to buy or not. So that is fourteen percent right but this is to be supported this result rule head is going to be supported that how often items in the rule body occur together as a percent of the total transaction so the result is fourteen percent right.

And that is how you are going to associate. So these are four elements which are very, very important that is Rule body, Confidence level, Rule head and Support in an association rule.

(Refer Slide Time: 13:06)



Now another thing is a market-based applications in market-based applications what happens you create scenarios so one example that have taken here is U.S. Company that is Petco what they have done they have basically used business analytics data basically and used Geographical information system and they basically whenever they want to locate a store, new store they want to open a new store so they have lot of factors to see that with the store is going to be profitable or not.

And what kind of risk is associated with that one so they have used this rule-based system. Now what they do basically they have software called ESRIGIS based site solution system which actually help them to mitigate the risk associated with networks of stores. Now see the risk increases as you increase the number of stores it also depends upon okay, whether your location what is your location of new store.

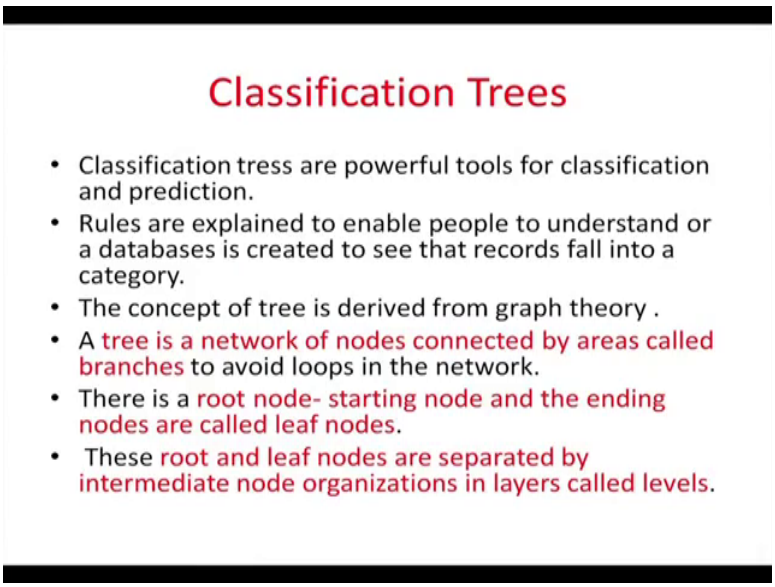
Because if your location of store does not have potential buyers okay then it is not going to mitigate the risk right, So they basically they use this kind of analysis to assess their sales potential for the new locations and what are the risk associated with this one and that is how they are able to successfully decide about the locations and the confidence level is very high because the use these kind of data to identify.

And based on that they are able to identify at least thousand two hundred locations

nationwide and they have been working successfully. Now apart from market-based analysis we also use classification trees these are very, very powerful tool for predictions okay and here rules are explained to people to understand or a database is created to see that database fall whether in this category or that category okay so basically this concept of trees has been derived from the graph theory.

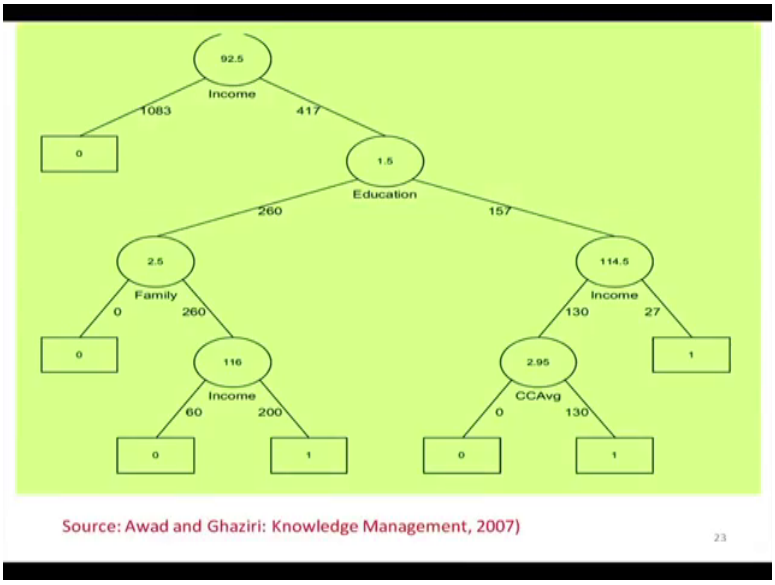
And tree is nothing but a network of nodes which are connected by area called branches to add to avoid loops in the networks so each nodes is connected and that is known as branches and there is a root node, starting node and then ending node are called leaf node right. And this root and leaf nodes are separated by intermediate node organizations or layers, So you can have different layers.

(Refer Slide Time: 15:24)



Now if you look at this example,

(Refer Slide Time: 15:25)

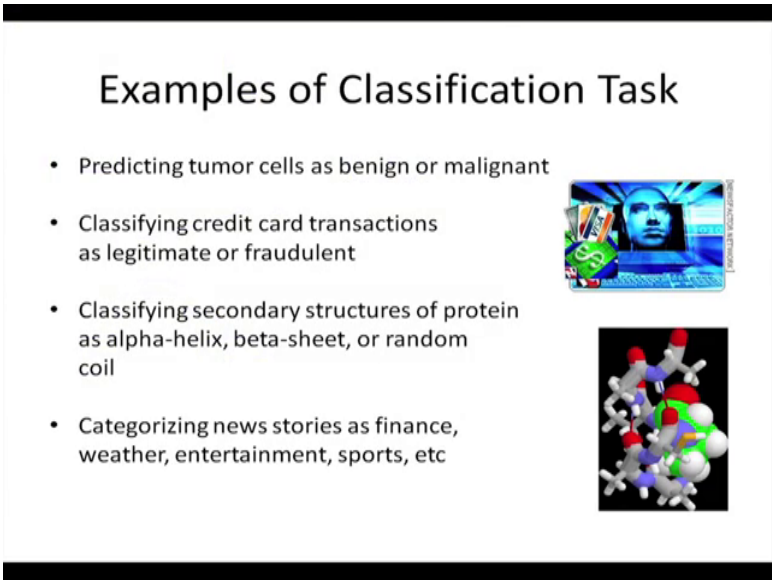


You can see this is level one this is layer two layer this is layer three and this is layer four right and that different variables based on that you are going to take different decisions right. The average the income is one variable zero or 1.5 so you gone for this then education that is 2.5 for this so there are certain variables based on that and there are is other variables family what is the family income then here income is what is average income.

And based on that what you are going to do. See tree this is the tree that we have formed then the root node, starting node, ending node this is the starting node that is the ending node. Now what we have then we have root leaf and root and leaf nodes are separate by intermediate node organizations. See if you look at intermediate node organization that these are intermediate node organizations which have been made there.

And then they are called leaf nodes okay so that is how we develop a tree this is also known as tree or decision trees.

(Refer Slide Time: 16:29)



Now let us take some examples like classification task that how you go about classifying activities for example based on certain rules you can see whether a particular tumor is benign or malignant right. So if you are going to take decisions and if you offer a person falls in that particular category based on records okay, then you can say the person has a tumor or of the tumor is benching so they need not worry or it is malignant.

And then this needs to be taken right for example you can classify credit card transactions whether they are legitimate or fraud right. Then you can also classify protein structures based on certain things or you can classify stories as finance weather, entertainment, sports right. For example new channels have certain categories as so this is how the class the tasks are classified and then you based on this you are going to develop classification trees okay.

And classify or predict an outcome based on a set of predictors you have these predictors and based on these predictors you are going to look at the outcome right.

(Refer Slide Time: 17:30)

Classification Trees: Rules

Goal: Classify or predict an outcome based on a set of predictors

The output is a set of **rules**

Example:

- Goal: classify a record as "will accept credit card offer" or "will not accept"
- Rule might be "IF (Income > 92.5) AND (Education < 1.5) AND (Family <= 2.5) THEN Class = 0 (non acceptor)
- Also called CART, Decision Trees, or just Trees
- Rules are represented by tree diagrams

For example you here if you look at this and the rule is that income and education and family if you look at this then whether it is going to be accepted or not accepted right. Will accept credit card or will not accept. So this decision to access are not accept is based on certain predictors and these predictors are income education and class right. So you are going to collect based on the data on this and you are going to make certain judgment okay.

This means advocate the income is above this, education is above this, the family size is this, they belong to a particular class then you are going to offer them credit card or you are going to not offer them credit card. So this is known as the decision trees are trees and the rules are printed by the diagrams know I have taken one example of a decision tree hear that you can look at for example if you see this okay.

(Refer Slide Time: 18:31)

Example of a Decision Tree

	categorical	categorical	continuous	class
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data

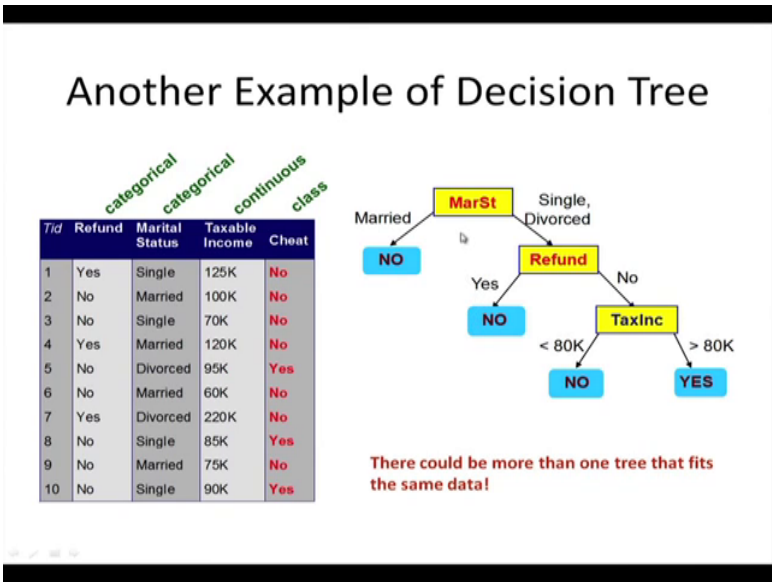
Model: Decision Tree

These are the things see categories okay, whether the refunds are made or not, this is the marital status and they to failures are whether the referrals are made of right, is the marital status and this is the taxable income were and whether they have cheated and not cheated right. So a splitting attributes are refund that is the first one so refund is No/Yes that you all can see. Refunded, No. Refund is there No.

And then you look at the marital status this is related to marital status and marital status whether you are single or divorced okay if you are married you are not even considered then you are going to see that whether tax what is the taxable income, whether it is less than eighty thousand or more than eighty thousand okay and then you have No or YES and this what you can see in Decision Tree right.

And this decision tree is based on this data which could be different sets of data because you have categories data you have continuous data because these are the continuous variable this is a nominal variable this is also a nominal variable and this is also a nominal variable class data okay whether you have done it or not done it.

(Refer Slide Time: 19:44)

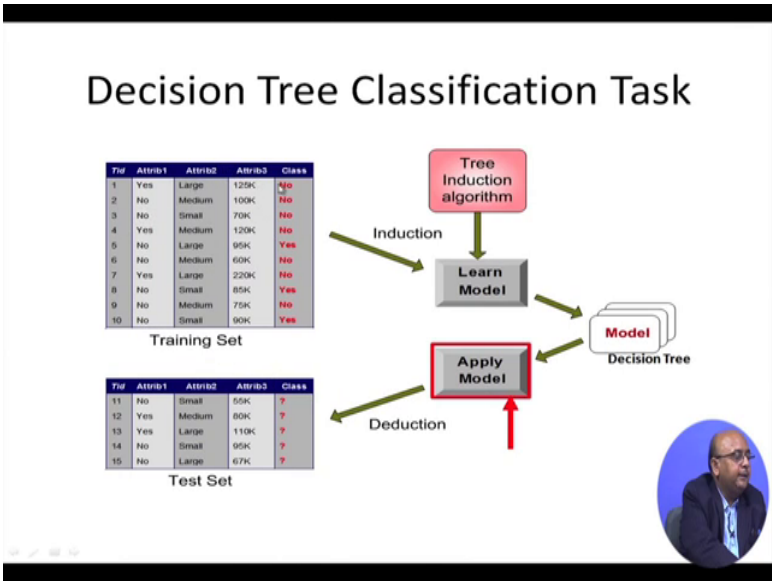


Another examples you can see here based on the same thing this what you have done you have changed the attributes right, here first of all you have gone for marital status single or divorced right, Refund yes/no then Taxable income Yes/No right and then you are going to see what is happening and then you are going to make this decision tree and this decision tree basically helps you to see that what is right.

And who is doing what so based on this you can say that who are the people, who are cheating right. For example you can see yes, those who are divorced have taxable income of eighty thousand and above right. Ninety five thousand and above are those who are single have an income of this are single not made a refund okay. But the income is ninety thousand yes they have cheated see you can identify see okay.

So based on that you can classify them into these are the people who have cheated and these are the people who have not cheated right. And then based on these the decisions would be based on that whether you are going to give them offer them credit card or not offer them credit card right.

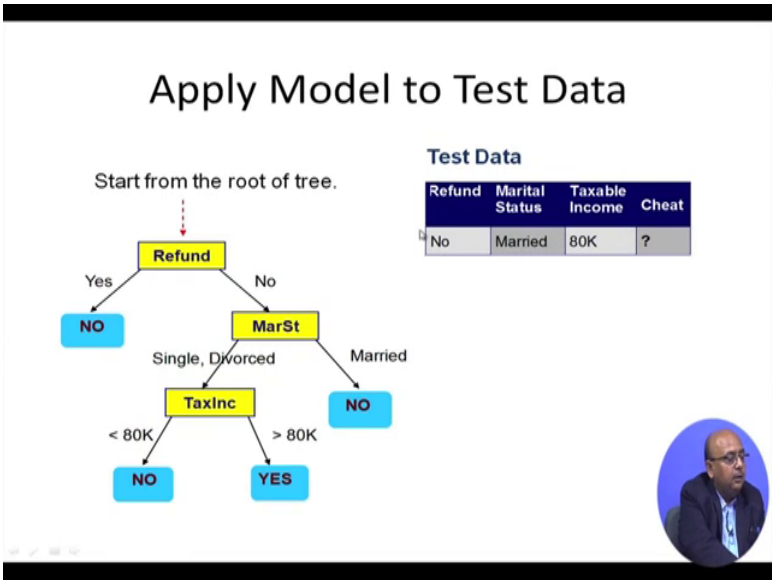
(Refer Slide Time: 20:57)



So basically based on this data that you have okay you are going to use them certain algorithm rules okay in the model and then you are going to apply this model it means that you are going to detect certain things out of it based on the data set okay so you are using these attributes 1, 2, 3 okay for example and then based on that you develop a decision tree and then you detect certain thing out of it that is the output okay.

So hear your going to induct certain so you are going to have a learning model and you are going to apply it and then based on that you have a model and which is going to help you to take a decision with what is to be done right.

(Refer Slide Time: 21:40)



Now look at another example these are test data okay which we have got. Refund no marital status yes married taxable income is whether he is going to cheat or not. Now how you are going to look at it. Start from the root of the tree refund – No, Marital status- Yes married No, Taxable income that is eighty thousand above so what you do see how the tree has been made. So based on that you can see whether this person is going to cheat or not.

So that include this person is not going to cheat okay. Now look into this look at another example you have these attributes yes/ no and these are for the different size large attributes, large, medium and small these are the income and these are the classes right. Based on that can we take a decision. Look at look at this so who is going to cheat, which class of people attribute one or no, so what you need to do you have to develop a decision tree.

And then you are going to see whether they are going to be classified as the cheater or not the cheater right . So based on this data you are going to form a decision tree and then you are going to see whether he is a cheater or not a cheater right, Thank you.