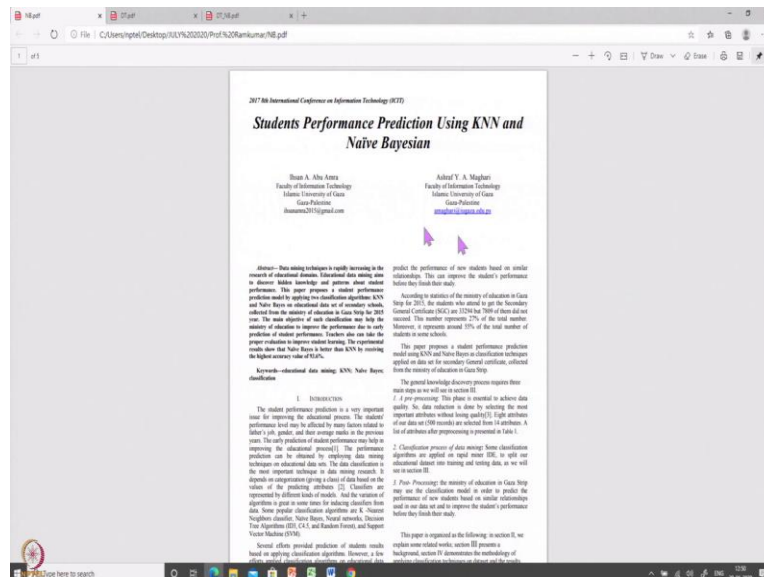


**Learning Analytics Tools**  
**Professor Ramkumar Rajendran**  
**Educational Technology**  
**Indian Institute of Technology, Bombay**  
**Lecture 9.5: DT, NB- Examples**

In this video we will see three different papers where decision tree and Naïve Bayes has been used. Since I was explaining the basics of decision tree and Naïve Bayes using, the simple example, let us look at the papers which used decision tree Naïve Bayes in research.

(Refer Slide Time: 0:35)



So let us look at Naïve Bayes first. So you know decision tree and Naïve Bayes has been used in research for long and there are lot and lot of publications on decision tree is available.

(Refer Slide Time: 0:54)



A man with dark hair, wearing a white and grey checkered button-down shirt, is shown from the chest up. He is looking slightly down and to his left, with his mouth open as if speaking. The background is a vibrant, abstract collage of various colors and patterns, including red, blue, yellow, and green, with some elements resembling faces or figures. The overall style is artistic and dynamic.

www.journals.tandf.co.uk

2017/7/1

ISSN 1043-9862 International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3, 2016

In this study, multiple classification techniques were used in the data mining process to produce the maximum grade at the end of the semester. This approach was used because it can provide a broader look, and understanding of the final results and output as well as, it will lead to a comprehensive conclusion over the outcomes of the study. Furthermore, a 10-fold cross validation was used to verify and validate the outcomes of the used algorithms and provide accuracy and precision measures.

(4) Data mining experimental and processing in this study was done using RapidMiner and WEKA.

As can be seen from Table 3 in the previous section (3.3), the results of the class attribute (GPA) in "Yes (Good)" which scores 10 times or 10% in the dataset. And hence, this percentage can be used as a reference in the accuracy measures provided by the algorithms in this research. Finally, in data mining, this is called the default model accuracy. The default model is a rule that predicts the class of all examples in a dataset as the class of the most frequent frequency. For example, in a sample dataset of 100 records and 7 classes (Yes & No), the "Yes" scores 70 times and "No" scores 30 times, the default model for this dataset will classify all objects as "Yes" hence, its accuracy will be 70%. Even though it is useless, the quality of output is a clue to evaluate the accuracy produced by other classification models. This concept can be generalized to all classification in the data to produce an expectation of the class result as well. Similarly, Table 4 was constructed to summarize the expected result for each class in the dataset.

**Table 3.3: Expected Results**

Class	Frequency	Percentage	Default	Yes	No
Yes (Good)	70	70%	Yes	70	30
No (Bad)	30	30%	No	30	70

**3.4. Decision Tree**

A decision tree is a supervised classification technique that builds a repetitive model from a given dataset attributes. The decision tree is a predictive modeling technique used for predicting, classifying, or categorizing a given data object based on the previously generated model using a training dataset with the same features (attributes). The structure of the generated model includes a root node, internal nodes, and leaf (terminal) nodes. The root node is the first node in the decision tree which has no incoming edges, and one or more outgoing edges. An internal node is a middle node in the decision tree which has one incoming edge, and one or more outgoing edges. The leaf node is the last node in the decision tree structure which represents the final suggested (predicted) class (label) of a data object.

In this study, four decision tree algorithms were used on the collected student's data, namely, C4.5 decision tree, ID3 decision tree, CART decision tree, and CHAID.

**C4.5 Decision Tree**

The C4.5 decision tree algorithm is an algorithm developed by Ross Quinlan, which uses the measure of the ID3 algorithm. The C4.5 algorithm uses pruning in the generation of a decision tree, where a node could be removed from the tree if such node was shown to be redundant.

Furthermore, the following settings were used with the C4.5 operator to produce the decision tree:

- Splitting criterion – information gain ratio
- Minimal size of split = 4
- Minimal leaf size = 1
- Minimal gain = 0.1
- Confidence = 0.5

After running the C4.5 decision tree algorithm with the 10-fold cross validation on the dataset, the following coefficient matrix was generated:

Actual \ Predicted		Class	
		Yes	No
Predicted	Yes	20	12
	No	10	20
Accuracy		0.67	0.67

The C4.5 algorithm was able to predict the class of 90 objects out of 270, which gives an accuracy value of 33.3%.

**ID3 Decision Tree**

The ID3 (Decision Information 3) decision tree algorithm is an algorithm developed by Ross Quinlan. The algorithm generates an improved ID3 algorithm from a dataset.

Following are the settings used with the ID3 operator to produce the decision tree:

- Splitting criterion – information gain ratio
- Minimal size of split = 4
- Minimal leaf size = 1
- Minimal gain = 0.1

After running the ID3 decision tree algorithm with the 10-fold cross validation on the dataset, the following coefficient matrix was generated:

Actual \ Predicted		Class	
		Yes	No
Predicted	Yes	20	12
	No	10	20
Accuracy		0.67	0.67

The ID3 algorithm was able to predict the class of 90 objects out of 270, which gives an accuracy value of 33.3%.

www.journals.tandf.co.uk

2017/7/1

ISSN 1043-9862 International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3, 2016

(Yes & No), the "Yes" scores 70 times and "No" scores 30 times, the default model for this dataset will classify all objects as "Yes" hence, its accuracy will be 70%. Even though it is useless, the quality of output is a clue to evaluate the accuracy produced by other classification models. This concept can be generalized to all classification in the data to produce an expectation of the class result as well. Similarly, Table 4 was constructed to summarize the expected result for each class in the dataset.

**Table 3.3: Expected Results**

Class	Frequency	Percentage	Default	Yes	No
Yes (Good)	70	70%	Yes	70	30
No (Bad)	30	30%	No	30	70

**3.4. Decision Tree**

A decision tree is a supervised classification technique that builds a repetitive model from a given dataset attributes. The decision tree is a predictive modeling technique used for predicting, classifying, or categorizing a given data object based on the previously generated model using a training dataset with the same features (attributes). The structure of the generated model includes a root node, internal nodes, and leaf (terminal) nodes. The root node is the first node in the decision tree which has no incoming edges, and one or more outgoing edges. An internal node is a middle node in the decision tree which has one incoming edge, and one or more outgoing edges. The leaf node is the last node in the decision tree structure which represents the final suggested (predicted) class (label) of a data object.

In this study, four decision tree algorithms were used on the collected student's data, namely, C4.5 decision tree, ID3 decision tree, CART decision tree, and CHAID.

**C4.5 Decision Tree**

The C4.5 decision tree algorithm is an algorithm developed by Ross Quinlan, which uses the measure of the ID3 algorithm. The C4.5 algorithm uses pruning in the generation of a decision tree, where a node could be removed from the tree if such node was shown to be redundant.

Following are the settings used with the C4.5 operator to produce the decision tree:

- Splitting criterion – information gain ratio
- Minimal size of split = 4
- Minimal leaf size = 1
- Minimal gain = 0.1
- Confidence = 0.5

After running the C4.5 decision tree algorithm with the 10-fold cross validation on the dataset, the following coefficient matrix was generated:

Actual \ Predicted		Class	
		Yes	No
Predicted	Yes	20	12
	No	10	20
Accuracy		0.67	0.67

The C4.5 algorithm was able to predict the class of 90 objects out of 270, which gives an accuracy value of 33.3%.

**ID3 Decision Tree**

The ID3 (Decision Information 3) decision tree algorithm is an algorithm developed by Ross Quinlan. The algorithm generates an improved ID3 algorithm from a dataset.

Following are the settings used with the ID3 operator to produce the decision tree:

- Splitting criterion – information gain ratio
- Minimal size of split = 4
- Minimal leaf size = 1
- Minimal gain = 0.1

After running the ID3 decision tree algorithm with the 10-fold cross validation on the dataset, the following coefficient matrix was generated:

Actual \ Predicted		Class	
		Yes	No
Predicted	Yes	20	12
	No	10	20
Accuracy		0.67	0.67

The ID3 algorithm was able to predict the class of 90 objects out of 270, which gives an accuracy value of 33.3%.

**CART Decision Tree**

Classification and Regression Tree (CART) is another decision tree algorithm which uses minimal cost-complexity pruning.

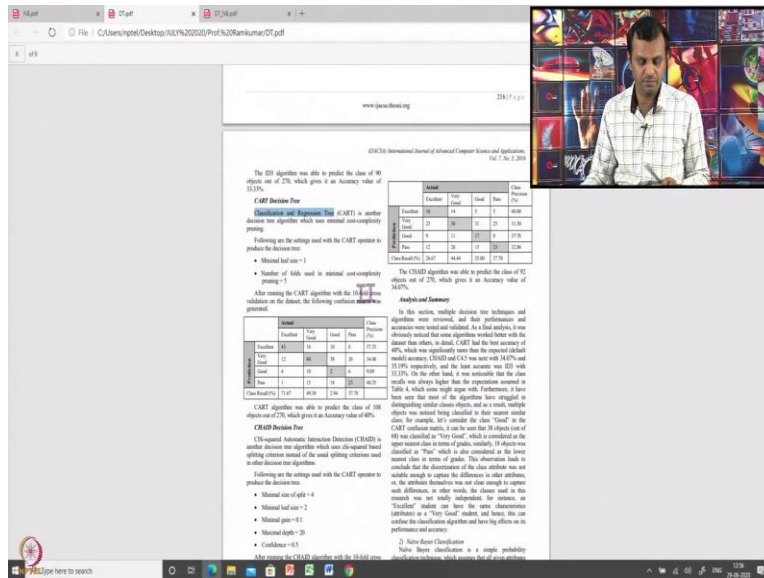
Following are the settings used with the CART operator to produce the decision tree:

- Splitting criterion – information gain ratio
- Minimal size of split = 4
- Minimal leaf size = 1
- Minimal gain = 0.1

After running the CART decision tree algorithm with the 10-fold cross validation on the dataset, the following coefficient matrix was generated:

Actual \ Predicted		Class	
		Yes	No
Predicted	Yes	20	12
	No	10	20
Accuracy		0.67	0.67

The CART algorithm was able to predict the class of 90 objects out of 270, which gives an accuracy value of 33.3%.



Let us look at the next paper that is for decision tree. In this paper they used decision tree to predict the student's performance. So, the data they used is the nationality, student gender, the first language. So what are the variables that is decision variables, category variables in decision tree.

So they have a multiple decisions to make and there are like a 1, 2, 3, 4, 5, 6, 7, 8 different features right? Or more than that see. So previous semester marks, we have week's friends and the father's occupation, mother's occupation, qualification of the father, the student discounts, any parents working in the university, all this information has been used. And they use this data to come up with a final dataset.

This is a final data set with the all the decisions and also they given the values in each data set like how they distributed across these particular values has been given. So, given this data set, they want to predict student's performance. So, this is just a distribution of the values, they want to show how the data is distributed, this is descriptive analytics. And given the table they want to predict students mark in the current semester.

They used to C4.5 decision tree algorithm with a ten fold cross validation on datasets, the confusion matrix is shown here. Now, you know what is cross validation and what is confusion matrix. Because now you are using, we have seen what is confusion matrix for two or three variables, now you are able to see what is confusion matrix and you are able

to make sense out of it. Classification is good, 46 percent like precision but recall is less because there are lot of excellent has been classified as very good.

So in a C4.5 class algorithm what is a criteria to split the tree, this is information gain ratio. And maximum size of pitch should be 4, not more than that, that you can say. Minimum leaf size should be 1. And the depth is how far the tree can grow. So they do not want to grow tree a very big or something, they want to make the trees pruning, the confidence gives the pruning criteria also. So they used C4.5 also ID3, it is exactly same information gain, maximum size, minimal gain, also depth also can be given.

And that value is this. And CART is classification and regression tree, as I mentioned, decision tree also can be used as regression. That is what they did. So they applied decision tree, different decision algorithms and they (results) reported results here. So this paper will help you to understand, what are the features they collected, how they combine the features, how they made a decisions, how they use it to make a decision and what are the results to compare. So this is one paper you can read and check it out.



(Refer Slide Time: 6:02)

The screenshot displays a web browser window showing a PDF document of a research paper. At the top, the browser's address bar and tabs are visible. The paper's title, "PREDICTING DROPTOUT STUDENT: AN APPLICATION OF DATA MINING METHODS IN AN ONLINE EDUCATION PROGRAM", is prominently displayed. Below the title, the authors' names, "Emine Yildirim" and "Serdar Korkmaz", along with their affiliations, are listed. The journal information, "European Journal of Open, Distance and e-Learning", volume, issue, and year, is also provided. The abstract section follows, detailing the study's purpose, methodology, and findings. A small inset photograph in the bottom right corner of the document area shows a man in a white shirt looking down at a device or screen. The overall interface includes standard Windows taskbar elements at the bottom.


2009 were included in this study. The percentage of students who registered and completed the program (N=123) were 63.67% and dropped the program (N=67) were 36.33%.

Table 1 presents the percentages of all participants' demographic characteristics. The number of male students (70%) was greater than the number of female (30%) students, and the students' ages ranged from 20 to 55 with an average of 28. The majority of the online program students were undergraduate and graduate students (MS or PhD student) (90.7%). Nearly half of the students (49.7%) have full-time or part-time jobs and only a few (39.5%) have previously been in an online course.

**Table 1. The Demographic Characteristics of Participants**

Gender	# of registered participants	# of dropout participants	percentage of registered participants	percentage of dropout participants
Female	31	26	25.83	37.48
Male	89	41	74.17	62.52
Age				
20-29	59	45	82.50	65.22
30-39	15	19	12.50	27.54
40+	6	5	5.00	7.25
Educational Level				
Graduate	45	30	37.50	43.48
Undergraduate student	59	32	49.17	46.50
Graduate student (MS or PhD student)	16	7	13.33	10.34
Previous Online Experience				
Yes	14	6	11.67	8.76
No	75	41	68.33	59.25
Occupation				
Working	59	35	49.17	50.72
Not working	31	34	30.83	49.29

*Eurasian Journal of Open, Distance and e-Learning - Vol. 17 / No. 1*  
 ISSN 2007-1059  
 © 2014 EJOEL



**Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program**  
*Emine Yildirim et al.*

Table 2 shows descriptive statistics of the participants' initial perceptions about online technologies self-efficacy, online learning readiness, levels of content, and prior knowledge at the beginning of the online program. Both registered and dropout participants had a positive perception about online activities based on their self-efficacy, skills, prior experiences



technologies self-efficacy, online learning readiness, locus of control, and prior knowledge at the beginning of the online program. Both registered and dropout participants had a positive perception about online technologies based on their self-efficacy scores. Most participants thought that they were ready for online learning, and also participants' perceptions in terms of their locus of control were generally low. It means that participants had better control of their behavior and perception for their own life and actions. Moreover, most participants stated that they did not have much knowledge about program courses at the beginning of the online certificate program. Finally, according to the results, the registered participants mean scores were slightly higher than dropout ones (see Table 2).

**Table 2** Descriptive Statistics of the Participants' Initial Perceptions about Predictors

	registered participants		dropout participants	
	Mean	Std.	Mean	Std.
online technologies self-efficacy (out of 100)	105.5	12.2	103.9	15.5
online learning readiness (out of 100)	41.26	5.91	43.7	4.5
locus of control (out of 20)	8.39	3.9	7.4	4.0
prior knowledge (out of 30)	13.13	4.89	12.3	4.3

**Data collection**

In this study, five online questionnaires were administered in order to determine participants' characteristics and initial perceptions: Demographic Survey (DS), Online Technologies Self-Efficacy Scale (OTSE), Readiness for Online Learning Questionnaire (ROLQ), Locus of Control Scale (LCS), and Prior Knowledge Questionnaire (PKQ). These selected questionnaires were translated into Turkish and used in previous studies. At the beginning of the online program, a five-hour face-to-face orientation was organized to explain the program and answer to the participants, help them meet each other and with the instructors, explain how to use the web page, and give information about the questionnaires related to the study. After the orientation, all online questionnaires were administered during the first week of the program. Participants submitted their responses to these online questionnaires and their data was stored on the database server.

DS was used to gather students' demographic information (e.g. age, gender, education level). OTSE was used to measure students' self-efficacy beliefs specific to the online environment. It was originally developed by Haidichar and Yi (2009) and it is a 24-item Likert scale with five subscales. This scale was translated into Turkish with a high Cronbach alpha level of 0.97 by one of the researchers of this study (Yakutlu, 2009).

ROLQ was used to assess students' readiness for online learning. It was originally developed by Nakita (2006) and composed of 13 items, rated by respondents on a 5-point scale. This scale was also translated into Turkish with a Cronbach alpha level of 0.76 by Yakutlu (2009).

European Journal of Open, Distance and e-Learning – Vol. 17 / No. 1  
July 2014



Haidichar and Yoon (2009) stress that DT is the most common DM technique in the literature. There are several popular decision tree algorithms such as ID3, C4.5, and CART (classification and regression trees). DT is in the form of a tree structure, where each node is either a leaf node (indicating the value of the target class of example) or a decision node (specifying a test to be carried out on a single attribute value, with one branch and sub-tree for each possible outcome of the test). Breiman, Friedman, Olsh, and Shaprio (1984) have many advantages such as very fast classification of unknown records, easy interpretation of small-sized trees, robust structure to the outliers' effects, and a clear indication of most important fields for prediction but DTs are very sensitive to over-fitting particularly in small datasets (Haidichar & Yoon, 2009).

In this study, to generate a decision tree, the C4.5 (Quinlan, 1993) algorithm was used, which is an extension of Quinlan's earlier ID3 algorithm. To construct the tree, entropy measure was used in the determination of nodes. Since the attributes with higher the entropy cause more uncertainty in outcome, they were selected in order of decreasing entropy.

**Naïve Bayes classifier (NB)**

A simple probabilistic classifier called as Naïve Bayes classifier was also used to conduct dropout classification. Naïve Bayes algorithm is the simplest form of Bayesian network (Domingos & Elkan, 1997) is one of the easiest algorithms to perform and has very satisfactory accuracy and economy rates (Kassambara, Perlembas & Pradisa, 2005). The posterior probability of each class,  $C_i$ , is obtained by the Naïve Bayes classifier using Bayes rule. The classifier makes the simplifying assumption that the attributes,  $A_i$ , are independent given the class, so the likelihood can be obtained by the product of the individual conditional probabilities of each attribute given the class (Flach & Lachiche, 2006). Thus, the posterior probability,  $P(C_i|A_1, A_2, \dots, A_n)$ , can be given by the following equation (assumption):

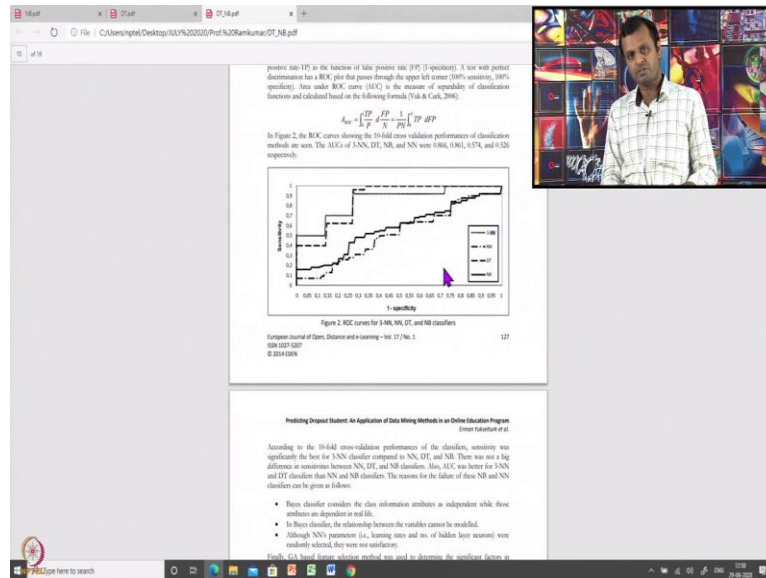
$$P(C_i|A_1, A_2, \dots, A_n) = P(C_i) \prod_{j=1}^n P(A_j|C_i) / P(A)$$

This assumption is usually called the Naïve Bayes assumption, and a Bayesian classifier using this assumption is called the Naïve Bayesian classifier often abbreviated as Naïve Bayes. Ultimately, it means that we are ignoring interactions between attributes within individuals of the same class (Flach & Lachiche, 2006).

European Journal of Open, Distance and e-Learning – Vol. 17 / No. 1  
July 2014

**Predicting Dropout Students: An Application of Data Mining Methods in an Online Education Program**  
Erman Yakutlu et al.





So let us look next paper. In this paper it has both decision tree and Naïve Bayes. So in this paper, it is the demographic information of the, like the participants like a female, male, age, graduate students and everything. And how many students registered, how many students dropped out. So let us see. So this, do you remember, this is the equation we saw in the Naïve Bayes classifier.

And this also explains how Naïve Bayes assumption is helping you to make this Bayes theorem and let us look at it. Yeah. I want to show this one thing. So they use the decision tree, Naïve Bayes, nearest neighbor, even neural network, do not worry about that. So Naïve Bayes is performed like this. Hope you understand this ROC curve. So Naïve Bayes is not good. Also the nearest neighbor is not different, but decision tree performed well.

So, I just want you to check these papers, understand what is the paper data you used and what are the classifiers they use, how they report the results. I hope now you are able to understand the paper, which be detailed because now you know what is algorithm, how does a matrix, so that helps you to understand the paper and which gives you identify the gap in the existing literature. And that might motivate you to collect data on your own and you can write your own paper, do some own research.

(Refer Slide Time: 7:50)

## Decision Tree Algorithm

- Saa, Amjad Abu. "Educational data mining & students' performance prediction." *International Journal of Advanced Computer Science and Applications* 7.5 (2016): 212-220.



## Naïve Bayes

- Amra, Ihsan A. Abu, and Ashraf YA Maghari. "Students performance prediction using KNN and Naïve Bayesian." *2017 8th International Conference on Information Technology (ICIT)*. IEEE, 2017.



## Decision Tree, NB

- Yukselturk, Erman, Serhat Ozekes, and Yalın Kılıç Türel. "Predicting dropout student: an application of data mining methods in an online education program." *European Journal of Open, Distance and e-learning* 17.1 (2014): 118-133.



So that three papers are given in this, in this slide, you can download and you can check those paper datas from scholar dot Google dot com. It is all available for you, if not put that in the forum. We can give a link where you can download that. So this three papers.

(Refer Slide Time: 8:08)

## Activity

### DT and NB

- Can you list down application of DT and NB and what data are required!

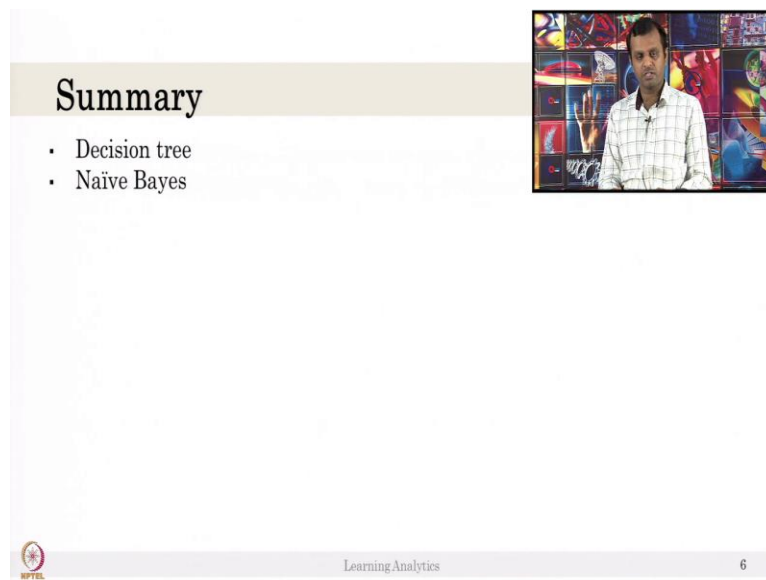


So, can you list down application of decision tree and Naïve Bayes, and also what data is required, what is the features required. And what are the categories and the features like what are the decision makings values and features. List down, this is based on, what is

decision tree now, what is Naïve Bayes now and you have seen three applications of papers.

Can you list down the applications of decision tree in learning analytics or the data from the learning environment? So, there is no answer. It is just if you list it down, just go ahead and try, try collecting data if possible and apply these algorithms. If this is a good and different than the existing research, please publish it in a good venue, international conference or journals.

(Refer Slide Time: 8:38)



**Summary**

- Decision tree
- Naïve Bayes

Learning Analytics 6

So in this week, we saw what is decision tree, what is Naïve Bayes. You understood what is Naïve Bayes and I talked about state transition sometime back in the diagnostic analytics. I would request you go and check Hidden Markov Model. If you know both, it is very easy to understand and very intuitive to go and next step will be that.

But that is not part of this course. That is completely optional. So I hope you understood what is decision tree and Naïve Bayes as usual. If you do not understand decision tree and Naïve Bayes, go and check the videos available internet. The idea here is to you to understand or is a concept, or logic, intuition begin these two classifiers, not to understand all the mathematical or training parameters to be used in the classifiers. Thank you.