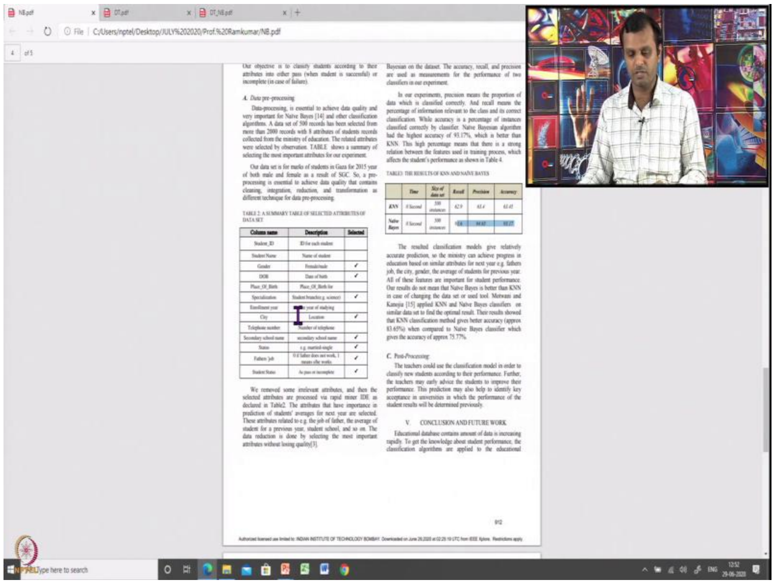
**Learning Analytics Tools**
**Professor Ramkumar Rajendran**
**Educational Technology**
**Indian Institute of Technology, Bombay**
**Lecture 9.5: DT, NB- Examples**

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)

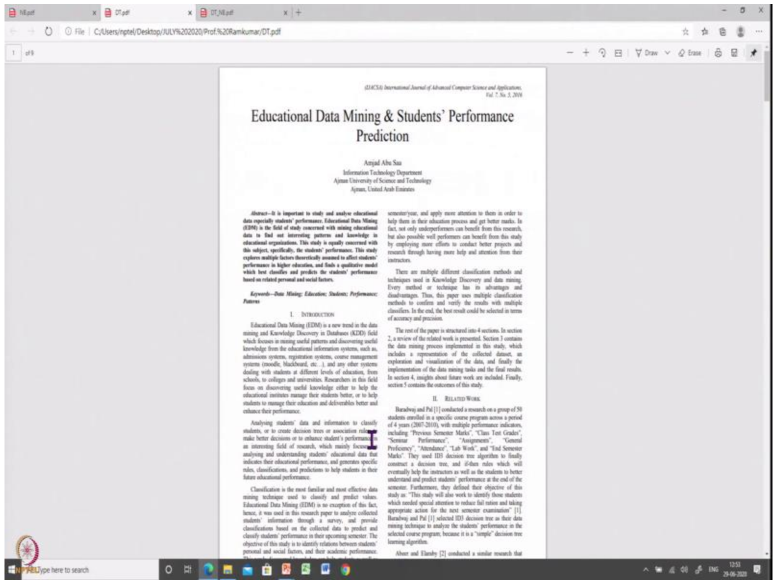


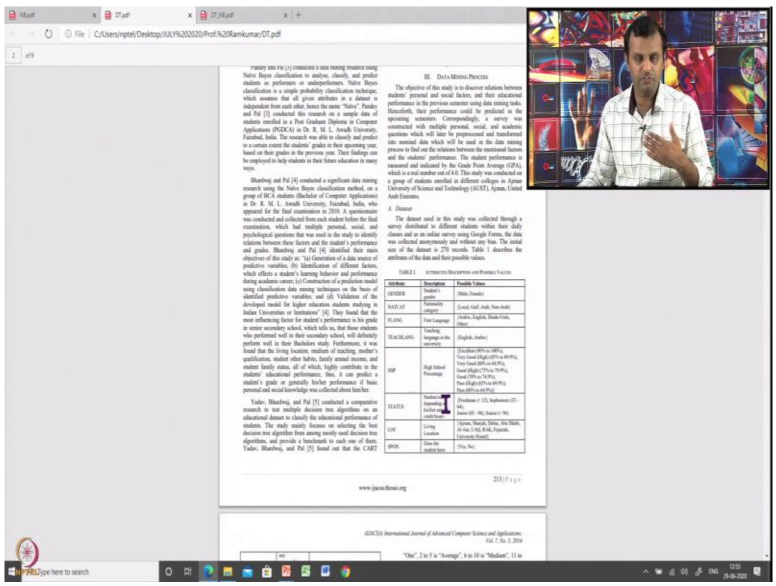
What is the data they considered to predict the performance? So to predict the performance they consider gender, date of birth, specialization, which city you are from, secondary school name, status, father's job, student status: as a pass or incomplete that is the thing we have to predict student status now, and the student is married or single, that is the marital status and all the information is captured here.

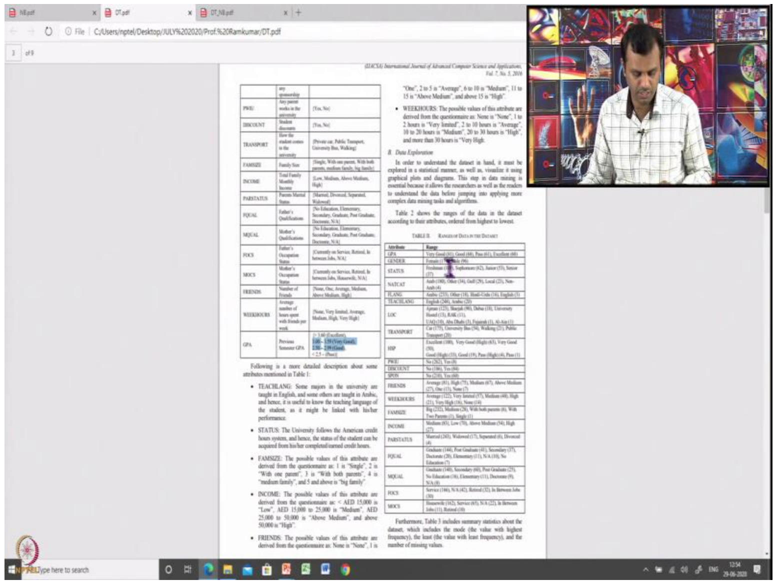
So, you have 1, 2, 3, 4, 5, 6, 7 features and one dependent variable to predict. So with these values and they computed a frequency table then they apply to K nearest neighbor and a decision tree. Now, in a decision tree,  it is accurate it is good, its performance is good that is the, that is a example, so it is very simple.

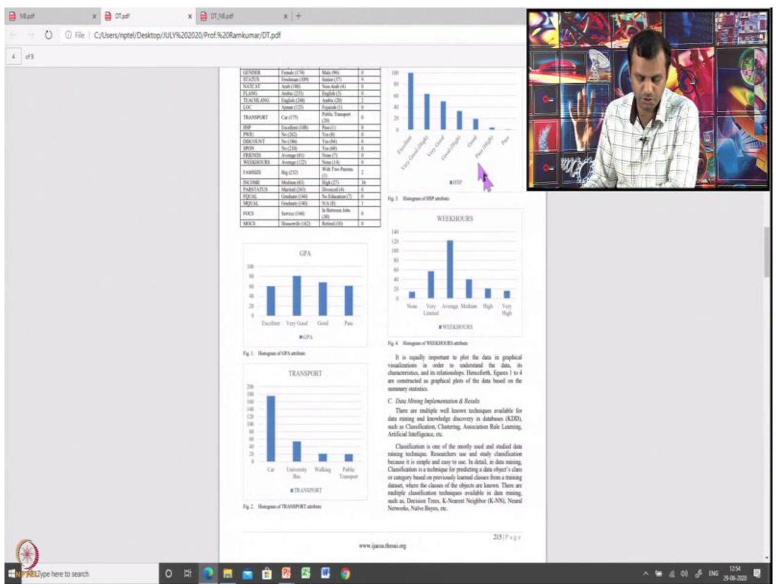
In this student's profile information is used and not their interaction with the system, not their attendance, or not the marks its or something, profile information is used to predict the students will pass the exam or not. Very simple paper, you can collect this information from survey questions and you can use that, right? So, you can check this paper, how they did it. They explained decision tree in a detail that is why I gave this paper, so just like, just check this paper.

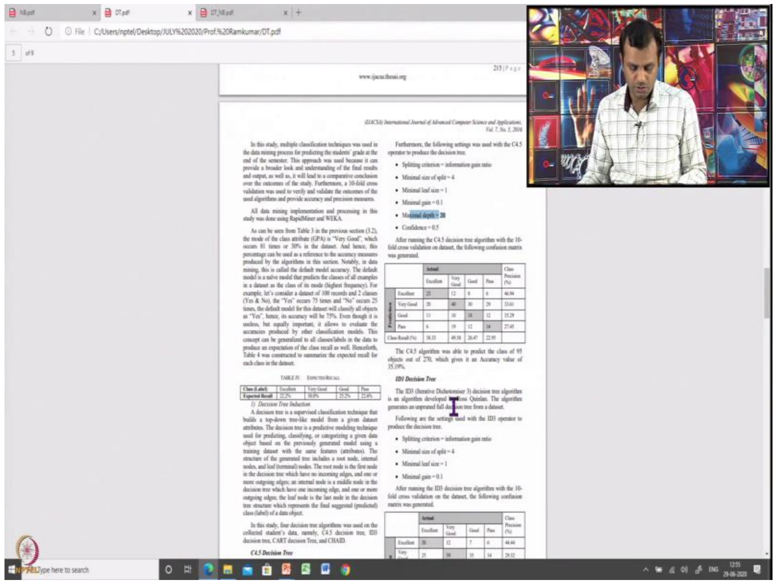
(Refer Slide Time: 2:22)

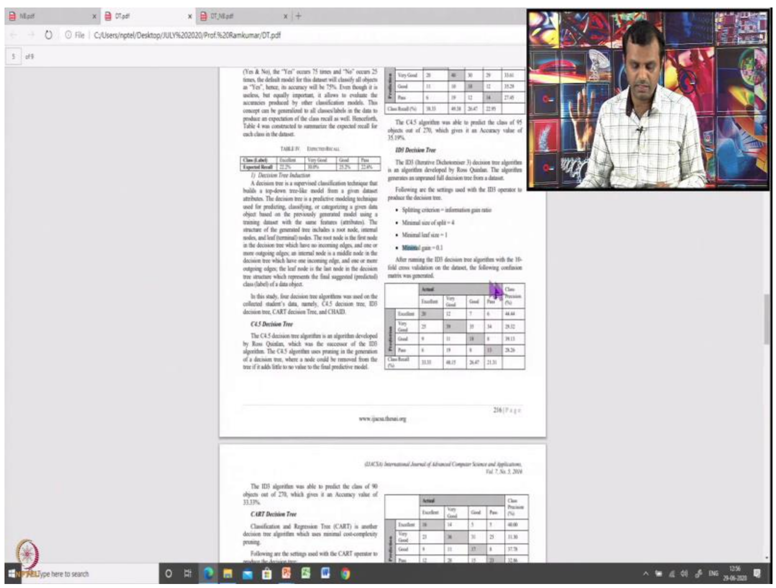

(IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 7, No. 5, 2016
Educational Data Mining & Students' Performance Prediction
Amjad Abu Saa
Information Technology Department
Ajman University of Science and Technology
Ajman, United Arab Emirates
I. Introduction
II. Related Work

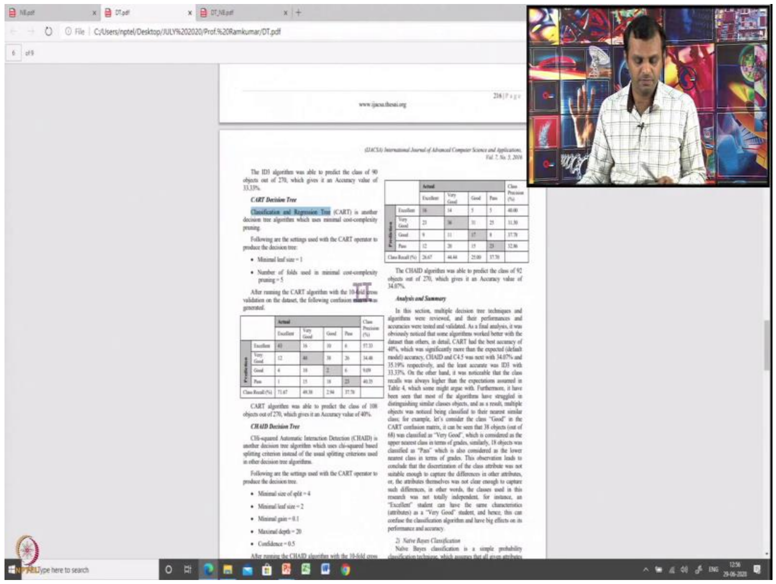

III. Data Mining Process
A. Dataset
Table I. Attributes Description and Possible Values
www.ijacsa.thesai.org




GPA
TRANSPORT
WEEKHOURS







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.

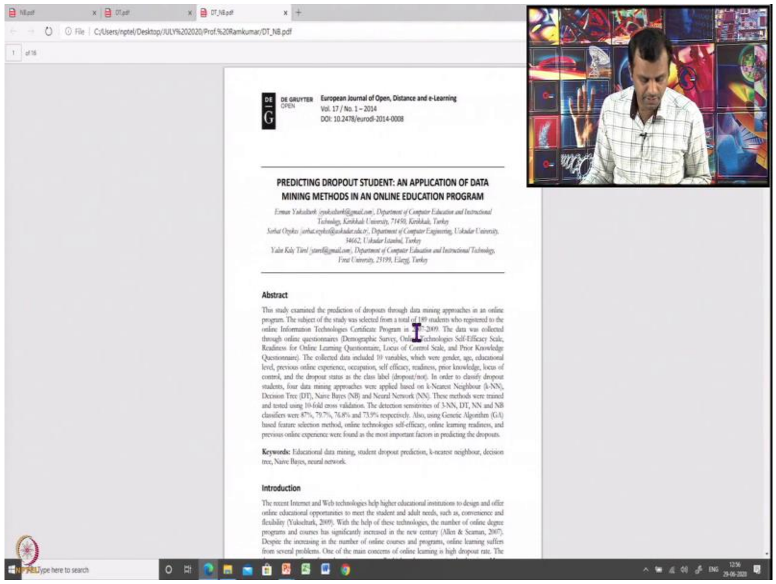
This is a final data set with the all the decisions and also they given the values in each data set like how they distributed across these particular values has been given. So, given this data set, they want to predict student's performance. So, this is just a distribution of the values, they want to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

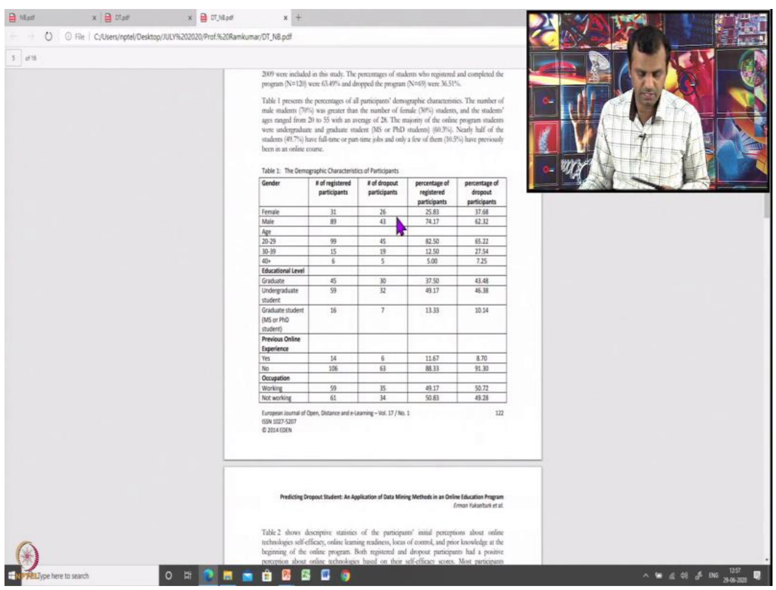
They used to C4.5 decision tree algorithm with a ten fold cross validation on datasets, the confusion matrix is shown here. Now, you know what is cross validation and what is confusion matrix. Because now you are using, we have seen what is confusion matrix for two or three variables, now you are able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

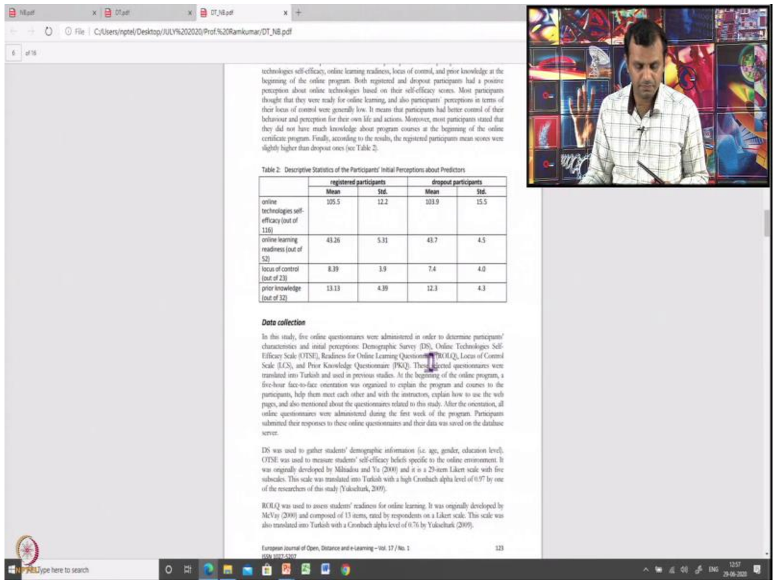
So in a C4.5 class algorithm what is a criteria to split the tree, this is information gain ratio. And maximum size of pitch should be 4, not more than that, that you can say. Minimum leaf size should be 1. And the depth is how far the tree can grow. So they do not want to grow tree a very big or something, 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)





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 behaviour 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 116) | 105.5 | 12.2 | 103.9 | 15.5 |
| online learning readiness (out of 52) | 43.26 | 5.31 | 43.7 | 4.5 |
| locus of control (out of 23) | 8.39 | 3.9 | 7.4 | 4.0 |
| prior knowledge (out of 32) | 13.13 | 4.39 | 12.3 | 4.3 |

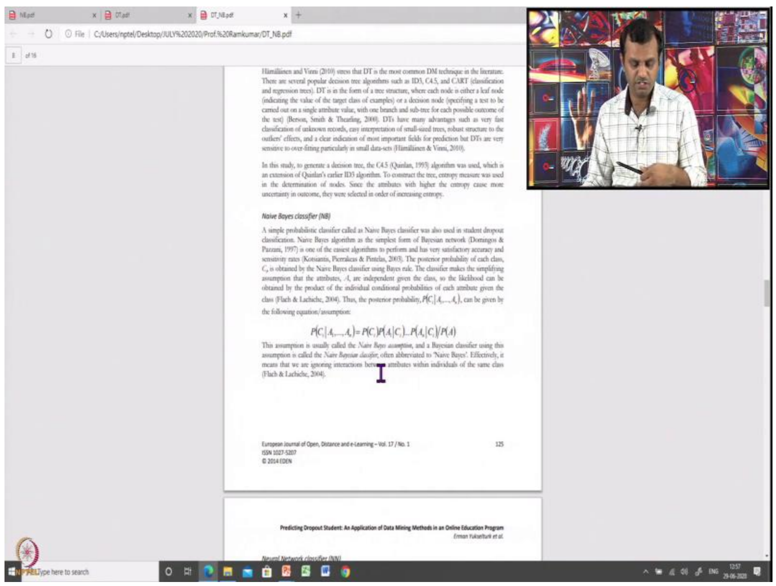
### Data collection

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 courses to the participants, help them meet each other and with the instructors, explain how to use the web pages, and also mentioned about the questionnaires related to this 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 saved on the database server.

DS was used to gather students' demographic information (i.e. age, gender, education level). OTSE was used to measure students' self-efficacy beliefs specific to the online environment. It was originally developed by Miltiadou and Yu (2000) and it is a 29-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 (Yukselturk, 2009).

ROLQ was used to assess students' readiness for online learning. It was originally developed by McVay (2000) and composed of 13 items, rated by respondents on a Likert scale. This scale was also translated into Turkish with a Cronbach alpha level of 0.76 by Yukselturk (2009).


ISSN 1027-5207



Hämäläinen and Vinni (2010) 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 examples) 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) (Berson, Smith & Thearling, 2000). DTs 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 data-sets (Hämäläinen & Vinni, 2010).

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 increasing entropy.

#### *Naive Bayes classifier (NB)*

A simple probabilistic classifier called as Naive Bayes classifier was also used in student dropout classification. Naive Bayes algorithm as the simplest form of Bayesian network (Domingos & Pazzani, 1997) is one of the easiest algorithms to perform and has very satisfactory accuracy and sensitivity rates (Kotsiantis, Pierrakeas & Pintelas, 2003). The posterior probability of each class, $C_i$, is obtained by the Naive 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, 2004). Thus, the posterior probability, $P(C_i|A_1,\ldots,A_n)$, can be given by the following equation/assumption:

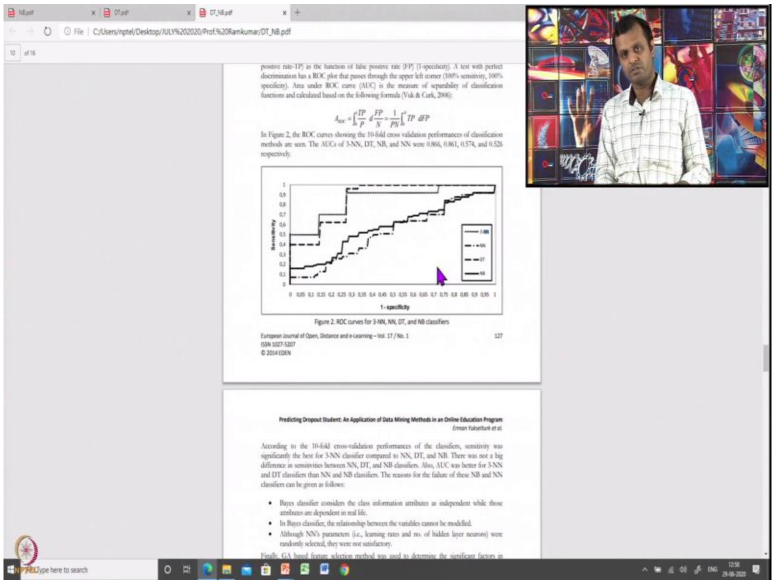
$$P(C_i|A_1,\ldots,A_n) = P(C_i)P(A_1|C_i)\ldots P(A_n|C_i)/P(A)$$

This assumption is usually called the *Naive Bayes assumption*, and a Bayesian classifier using this assumption is called the *Naive Bayesian classifier*, often abbreviated to 'Naive Bayes'. Effectively, it means that we are ignoring interactions between attributes within individuals of the same class (Flach & Lachiche, 2004).


ISSN 1027-5207




#### *Neural Network classifier (NN)*



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)

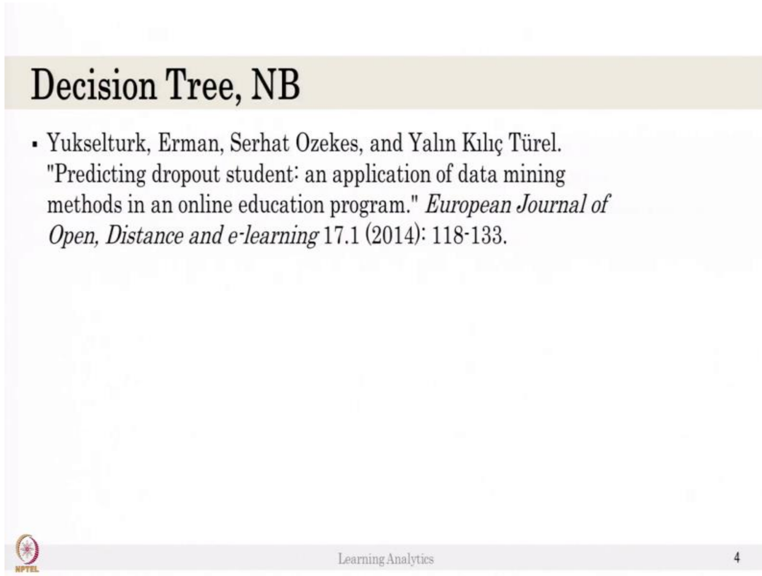
# Decision Tree Algorithm

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



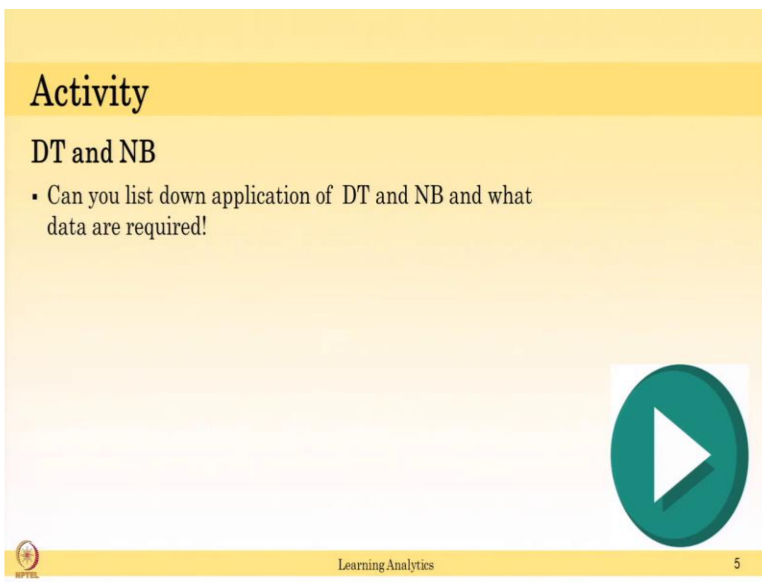
# Naïve Bayes

- 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.




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.

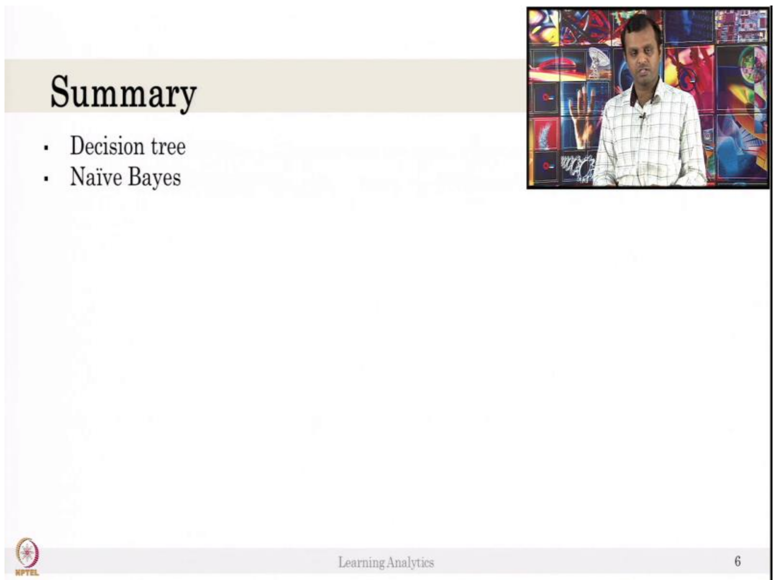
(Refer Slide Time: 8:08)




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)




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.