



**IIT Madras**  
ONLINE DEGREE

**Computational Thinking**  
**Professor. Madhavan Mukund**  
**Department of Computer Science**  
**Chennai Mathematical Institute**  
**Professor. G. Venkatesh**  
**Indian Institute of Technology, Madras**  
**Top-down vs. Bottom-up approach and concept of decision tree**

**G Venkatesh:** So far, the entire lectures that we had so far, all the problems that we took, had a very top-down structure, we started with the problem definition. And then we tried to find a way that would work for all the possible cases, right?

**Madhavan Mukund:** Right.

**G Venkatesh:** So, and then slowly, we developed it step by step, we develop the algorithm, and then, regardless of what data we give it, the algorithm should work, that was how it was constructed.

**Madhavan Mukund:** So, we will have a uniform algorithm,

**G Venkatesh:** We will have one algorithm or one method,

**Madhavan Mukund:** Which is based on some problem analysis that we have,

**G Venkatesh:** Analysis of the problem,

**Madhavan Mukund:** And this will always work for an algorithm,

**G Venkatesh:** And to make an algorithm, it will be an algorithm fix the structure of,

**Madhavan Mukund:** Yeah, some steps,

**G Venkatesh:** Steps and it is fixed, right? rigid way, in some sense. There is a flowchart or a pseudo-code or something, which is written down, and then you are not allowed to change it. Right?

**Madhavan Mukund:** Correct.

**G Venkatesh:** And then as the data comes, the program keeps executing on the data and gives you a result. And you might rewrite the program and create another program, but the program itself is not changing,...

**Madhavan Mukund:** Correct.

**G Venkatesh:** Simply because it saw a new data item.

**Madhavan Mukund:** Correct.

**G Venkatesh:** Right. Program is not changed, that is not how real life works.

**Madhavan Mukund:** that is true. I mean, sometimes when we make decisions, right, so we look at various things, and we decided to do one thing or another thing, for instance, we might be waiting at a bus stop. And we might want to decide whether to give up and catch the next bus. I mean, maybe the bus is not going all the way. But you have to change buses somewhere? Is that better or wait for the arrival of another bus?

**G Venkatesh:** Wait for the another...

**Madhavan Mukund:** So...

**G Venkatesh:** Or take an auto.

**Madhavan Mukund:** Yeah, so some sort such thing. So, we cannot. I mean, what works today may not work tomorrow, or over a period of time the bus frequency may change.

**G Venkatesh:** Correct,...

**Madhavan Mukund:** So, we might have to change our strategy.

**G Venkatesh:** And you know, if you generally ask somebody is the bus likely to come, the guy will say, I do not know, I am the first time, I am also here first time.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Or he may say no, no...

**Madhavan Mukund:** Or somebody...

**G Venkatesh:** Who is already...

**Madhavan Mukund:** The tea shop guy might say that the bus just went, maybe it came early today. So, then it is not.

**G Venkatesh:** After one month, if you keep standing every day and getting to the bus after one month, you will figure it out.

**Madhavan Mukund:** correct.

**G Venkatesh:** You learn.

**Madhavan Mukund:** So, you learn the learner strategy by watching,...

**G Venkatesh:** you learn by watching, experiencing, but we are not doing all that?

**Madhavan Mukund:** So far, we are not doing anything like that.

**G Venkatesh:** So, what we learning to experience would be some kind of a bottom-up.

**Madhavan Mukund:** Yes, instead of taking a strategy and applying it to all the data, you are taking the data and deriving a strategy from it. So...

**G Venkatesh:** Incrementally,...

**Madhavan Mukund:** Incrementally.

**G Venkatesh:** As the data comes,...

**Madhavan Mukund:** As you see more and more data, you literally have to adapt your strategy to make it. But of course, this strategy is not going to be 100 percent. That that is the nature of me the nature of the problem is also likely?

**G Venkatesh:** Because the problem is,...

**Madhavan Mukund:** Because,...

**G Venkatesh:** Not all problems.

**Madhavan Mukund:** we will have this...

**G Venkatesh:** kind of thing is good fun. So, the problem is pretty structured, very well defined. A top down might work.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Problem is a little bit fuzzy.

**Madhavan Mukund:** Yeah, there you have to make a call on,

**G Venkatesh:** And it may also sometimes think were working once, for one set of data may not work for another set, but you are okay. Because this is a set of data you are likely to encounter. Anyway,...

**Madhavan Mukund:** Correct.

**G Venkatesh:** Isn't it?

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So suppose I say that this thing is not going to work for a very large number of students in the class, it will work only if there are only 60 students in the class. I said quite okay, that I have only 60 students in my class, I do not want it to work for 6000 students and 600,000 students, isn't it?

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So sometimes, if it works for my class, and I, my class is going to be constant for the whole year, figured out it works for my class, it works for my class, at least I can use it for this year, next year. Maybe I have tried something else we will see next year.

**Madhavan Mukund:** Right.

**G VENKATESH:** So, this bottom-up seems to be a new method. But,...

**Madhavan Mukund:** As you do not go about it, that is not good.

**G Venkatesh:** Take some example problems. Now, I was thinking this to take this classroom data set, right? I was looking through it when we are doing all these problems, I am working through look like the, from the marks or something. You can make out a student test Suppose I do not tell you who the student is. I do not say this is data. I just tell you the total is 173. Physics marks is 42. Let us say or something is marked 53. Of these two, can you identify a student?

**Madhavan Mukund:** okay, so you want to identify the student from some information, but not the name.

**G Venkatesh:** No.

**Madhavan Mukund:** Okay.

**G Venkatesh:** It is like, you know, in face recognition, they just look at the nose and figure it out, something like that...

**Madhavan Mukund:** Correct. Correct.

**G Venkatesh:** So, our fingerprint. We can have ...

**Madhavan Mukund:** Correct.

**G Venkatesh:** You can have....

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Almost. Exactly.

**Madhavan Mukund:** Yes.

**G Venkatesh:** So, Aadhar works.

**Madhavan Mukund:** And they measure all these distances. And

**G Venkatesh:** Something they do..

**Madhavan Mukund:** So, there is some numbers which come in, and then they figured out from that, who it is?

**G Venkatesh:** Who it is?

**Madhavan Mukund:** So, in this way, we have some numbers. And we want to know if the numbers determine the student.

**G Venkatesh:** Yeah.

**Madhavan Mukund:** Rather than the other way or...

**G Venkatesh:** The other way around it.

**Madhavan Mukund:** So, we know which student has got the marks. But can we guess the notable student from the marks.

**G Venkatesh:** Marks.

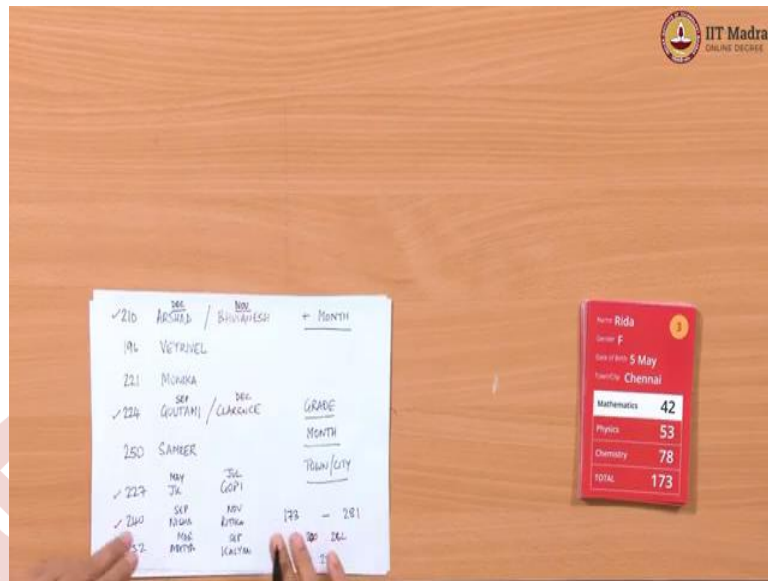
**Madhavan Mukund:** That is what we want....

**G Venkatesh:** So, let us start with something. Let us say we have the total.

**Madhavan Mukund:** G Venkatesh. So, we want to say that if we know the total, we know the student,...

**G Venkatesh:** Correct. So, we can take so randomly pick one card. 210.

(Refer Slide Time: 05:27)



**G Venkatesh:** Right.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Now, we have to check. There may be other students with 210. We do not know.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** For now, I can note down...

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So, 210 is Arshad.

**Madhavan Mukund:** So, it should be noted it down,

**G Venkatesh:** You can note it down.

**Madhavan Mukund:** Okay.

**G Venkatesh:** I can say like this, and randomly take another card. As long as this works this fine.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Take another card randomly.

**Madhavan Mukund:** So, 196



**G Venkatesh:** Vetrivel 196. So far, no clash.

**Madhavan Mukund:** Okay.

**G Venkatesh:** So total, I am able to guess the student, it seems two examples.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Okay, the third one, Monica 221.

**Madhavan Mukund:** Okay,

**G Venkatesh:** So, you tell, 221 I will say, Monica.

**Madhavan Mukund:** So, so far, it is looking good. So, if you give me a number, and if I know this table, then I can look up the table and say, if it is 211, it is Monica

**G Venkatesh:** Monica,

**Madhavan Mukund:** Of course. How many things do I need before I can agree on this? So, shall we try this now? So supposedly, can you conclude so there are about 30 students in this class? So, we picked up some three random ones, and they turned out to have different totals.

**G Venkatesh:** Correct

**Madhavan Mukund:** So, is this a good enough sample for us?

**G Venkatesh:** I don't know. We should try a few more, let us try a few more.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Let us see random 224. It is close but different.

**Madhavan Mukund:** So, this is Goutami.

**G Venkatesh:** We try one more, Sameer 250.

**Madhavan Mukund:** Okay.

**G Venkatesh:** So, what seems to work? Now, of course, one can say we have got 6. So, 6 out of 30 is quite a big sample

**Madhavan Mukund:** 5 out of 30. 5 here

**G Venkatesh:** 41



**Madhavan Mukund:** Yeah

**G Venkatesh:** 1 by 6.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** You have got a normally,...

**Madhavan Mukund:** It is more than 15 percent.

**G Venkatesh:** So, it looks How do you say?

**Madhavan Mukund:** Well, one way to route is just to check I suppose, whether there are any students but which we can do in a small enough set. We can check whether there are two students with that same total.

**G Venkatesh:** Yeah.

**Madhavan Mukund:** In general, this may be hard to do if we do not have all the information.

**G Venkatesh:** Sorted.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Let us assorted it,...

**Madhavan Mukund:** Okay.

**G Venkatesh:** So, we just have to check it sorted here it says go and see whether the neighbouring cards

**Madhavan Mukund:** Yeah correct,

**G Venkatesh:** 181 187 we have taken out some

**Madhavan Mukund:** Yeah.

**G Venkatesh:** 188 189 plus.

**Madhavan Mukund:** 196 is here.

**G Venkatesh:** 198. So far, so far, so good.

**Madhavan Mukund:** 29 210 is here. So, we have..

**G Venkatesh:** So, by 210 you cannot say...

**Madhavan Mukund:** because 210. We have Arshad, we have already found and now we have

**G Venkatesh:** Bhuvanesh, so then what we do

**Madhavan Mukund:** Okay, so we say..

**G Venkatesh:** To guys 210.

**Madhavan Mukund:** So, this like a tie? Right? So, your tiebreaker.

**G Venkatesh:** Arshad. Tiebreaker Arshad Bhuvanesh. So, Arshad and Bhuvanesh. So, what to we do now.

**Madhavan Mukund:** So, let us see there are other things on this card. So of course, we have other marks, but we already have the totals. So maybe we should look at something which is not marks. So we can say maybe...

**G Venkatesh:** Month of birth.

**Madhavan Mukund:** Yeah. Which month they are born...

**G Venkatesh:** Date of Birth till obvious because everybody will have different date of birth...

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Actually, I should say but...

**Madhavan Mukund:** Yeah. So, let us take...

**G Venkatesh:** Date of birth.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So, you want to use minimal information. We can take a city, let us say a different city, we can take a month of birth, or we can take one mark also, we can take say physics marks...

**Madhavan Mukund:** So, which one should we do?

**G Venkatesh:** Let us start with month, birth, month of birth

**Madhavan Mukund:** So now we are saying that if they are equal on total, then we will check that they are not equal on month of birth...

**G Venkatesh:** Yeah.

**Madhavan Mukund:** So, it should be a different one.

**G Venkatesh:** Month of birth...

**Madhavan Mukund:** Month of birth Okay. So, we saw one duplicate just 210. So, if we continue, let us see if there are more duplicates and see whether this works. So, this so now we are saying plus.

**G Venkatesh:** So, so we are told in case of Arshad. So, we will be given basically two things in case Arshad will say December.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** 210 will say...

**Madhavan Mukund:** Yeah. Correct. So, I will just write it on top of December and I think Bhuvanesh was November.

**G Venkatesh:** This was November.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** And 216 No,...

**Madhavan Mukund:** No, no.

**G Venkatesh:** 219 224, uh 224 also Okay, Clarence and Goutami. All right. Now this is September this is December, so it does discriminate.

**Madhavan Mukund:** but already in those five that we got. We had already got two cases where they were not unique, which is quite surprising. September and December, okay. But at least with the month, it works.

**G Venkatesh:** The month it seems to work. So far it works....

**Madhavan Mukund:** 224.

**G Venkatesh:** 227.

**Madhavan Mukund:** 227, it will have....

**G Venkatesh:** Here it is 227.

**Madhavan Mukund:** Okay.

**G Venkatesh:** JK and Gopi.

**Madhavan Mukund:** But May

**G Venkatesh:** July.

**Madhavan Mukund:** so that works. So, should I write it down.

**G Venkatesh:** Yeah.

**Madhavan Mukund:** So, we have 227 JK, Gopi, here May and July. So so far this rule is working.

**G Venkatesh:** We will make it harder. I think I do not want to give them exact minus give exact minus 240.

**Madhavan Mukund:** 240 is also 240.

**G Venkatesh:** Nisha and Ritika, September and November, so, again, a month seems to work. So, 244 247 52 March and September.

**Madhavan Mukund:** Okay, so this is another duplicate. Aditya and Kalyan.

**G Venkatesh:** Then is 254 261 276 285. I think these are all different? No, let us say for a moment, I do not give you the exact answer because exact marks already have too much information. I know.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** I give you only a grade.

**Madhavan Mukund:** Correct. Let us say you get only the grade

**G Venkatesh:** A, B, C, D.

**Madhavan Mukund:** Okay.

**G Venkatesh:** So, then the value of total ability for the total to discriminate goes on the grading.

**Madhavan Mukund:** correct. So, whatever I use grade,

**G Venkatesh:** I have only the grade.

**Madhavan Mukund:** Grade as a total grade.

**G Venkatesh:** Total grade.

**Madhavan Mukund:** Okay.

**G Venkatesh:** And I have month.

**Madhavan Mukund:** Okay.

**G Venkatesh:** And let say I have town.

**Madhavan Mukund:** so these three are enough that we can try.

**G Venkatesh:** So, let me put this back so we can write the grades for the students. So, we have to decide the grade first. So, what grades do we give? We are given it earlier, I think, but we can do it again. We have divided the range and divide it into range.

**Madhavan Mukund:** Okay, so we will divide into quarters

**G Venkatesh:** Quarters. So, we are going from 173 281.

**Madhavan Mukund:** So, that is 108.

**G Venkatesh:** Divide by 4,

**Madhavan Mukund:** So, 27.

**G Venkatesh:** 27.

**Madhavan Mukund:** So, you want to do it according to equal Mark bond, so we will have said like 210 and 237. No, it does not make sense. Something's wrong 237 and then....

**G Venkatesh:** This is D, this is E this is B and that is a ...

**Madhavan Mukund:** so, I will have 27. So, I will have 264.

**G Venkatesh:** No, 173 281 is 100 and...

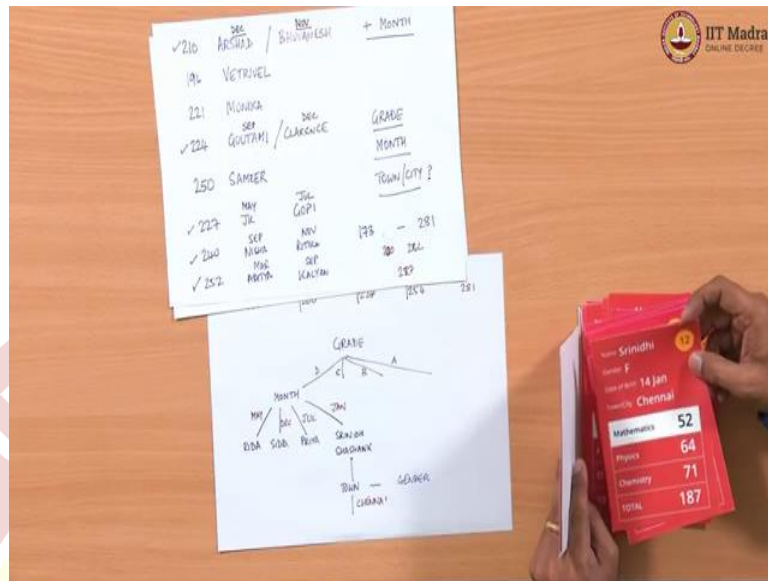
**Madhavan Mukund:** This will be 200 .

**G Venkatesh:** 54 27.

**Madhavan Mukund:** So, it will be 200 227.

**G Venkatesh:** 200 227 254.

(Refer Slide Time: 14:02)



**Madhavan Mukund:** Yeah, so let us write it down properly. So, the grades now are 173 200 227 254 and 281. So, everything below 200 will be D.

**G Venkatesh:** D.

**Madhavan Mukund:** Everything below 226. And above 200 will be a C grade.

**G Venkatesh:** C grade.

**Madhavan Mukund:** This will be B grade. And everything in this uppermost thing will be...

**G Venkatesh:** A grade.

**Madhavan Mukund:** A grade. So now,...

**G Venkatesh:** So, I am only told the grade,....

**Madhavan Mukund:** Yes, I am told the grade, which is in these...

**G Venkatesh:** Which means that these guys all these people from 173 up to 200.

**Madhavan Mukund:** Will have the same...

**G Venkatesh:** Will have the same grade so you cannot discriminate all of these now,...

**Madhavan Mukund:** on the grade. Now, I will ask you, give you the grade and then give you the month of birth.



**G Venkatesh:** Now I can tell you the month of birth,...

**Madhavan Mukund:** So now maybe we can just see in this thing, how many how they divided up according to month. So, we have May.

**G Venkatesh:** We have a table, now, this.

**Madhavan Mukund:** Okay, so we have...

**G Venkatesh:** So, we have first thing is your started with?

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Grade.

**Madhavan Mukund:** Grade. So, we start with grade. So, let us...

**G Venkatesh:** Make their, this is like decision tree or something.

**Madhavan Mukund:** So, we can take the grade. So, right?

**G Venkatesh:** First check the grade,...

**Madhavan Mukund:** And then in the grade, you can fall into these 4....

**G Venkatesh:** 4 buckets

**Madhavan Mukund:** 4 categories.

**G Venkatesh:** So, the first is a check the grade.

**Madhavan Mukund:** Yeah, so the four categories are D, C, B, and A.

**G Venkatesh:** A.

**Madhavan Mukund:** So now, we are going to check for each of these. We are going to check the month...

**G Venkatesh:** We will check the month. A month can be 12.

**Madhavan Mukund:** Yeah, but we will see because we will only record the months that we have, but there could be in principle 12.

**G Venkatesh:** 12 possible branches could be there.

**Madhavan Mukund:** Yeah.



**G Venkatesh:** So, this May Rida.

**Madhavan Mukund:** Okay.

**G Venkatesh:** If another person falls in May, that means this is not enough.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Right.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** We can put there that leaf as Rida

**Madhavan Mukund:** Should I just, so, I will write Rida Okay.

**G Venkatesh:** This one is December.

**Madhavan Mukund:** So, I will not write it in order illustrated as it helps that Siddharth

**G Venkatesh:** Then July is Priya. Then Srinidhi is Jan. So far, so good.

**Madhavan Mukund:** So far, they are all different. Yeah.

**G Venkatesh:** Jan another Jan.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Okay.

**Madhavan Mukund:** So, already.

**G Venkatesh:** Already, yeah, they are colliding. So, Shashank. So, the two

**Madhavan Mukund:** Were in January.

**G Venkatesh:** In January after discriminate between these two.

**Madhavan Mukund:** So, we said that we would go back town

**G Venkatesh:** Town is also Chennai.

**Madhavan Mukund:** So now we are in trouble. So, this is not going to work. So, if I say now town,...

**G Venkatesh:** Again, they will come to the same again.

**Madhavan Mukund:** So now I say town. They are both in Chennai. So, these three Grade, Month and Town City.

**G Venkatesh:** Are not enough.

**Madhavan Mukund:** Are actually not enough,...

**G Venkatesh:** But I have the gender, they are different in gender.

**Madhavan Mukund:** so different in gender. So how do we find this out, so we want some three things?

**G Venkatesh:** that is the question. I think, yeah, that is the key question that you are asking.

**Madhavan Mukund:** Or, instead we are saying, ask gender.

**G Venkatesh:** Ask gender.

**Madhavan Mukund:** Because that happened to work for these two, but now we may find that go to some other category. Grade.

**G Venkatesh:** These are, this is how learning works. Do you try it?

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Then keep changing.

**Madhavan Mukund:** Okay, so now we do gender,

**G Venkatesh:** Let us do gender.

**Madhavan Mukund:** So, let us do gender.

**G Venkatesh:** Now again, the question is whether it should start with Grade 1'st or you could start with something else, or we should start with Month first, because month is dividing more, in 12.

(Refer Slide Time: 17:49)



**Madhavan Mukund:** Month is divided into 12, genders only dividing into 2, so it is not going to be?

**G Venkatesh:** Not a good idea.

**Madhavan Mukund:** Good idea.

**G Venkatesh:** Start with month first and then Grade.

**Madhavan Mukund:** So, you start with month first.. So, then we have these 12 different things. So...

**G Venkatesh:** So, we will go one by one again.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Ideally one should take a sample and one how do we do this? Actually, typically, we will take a small subset and you do it see whether it works.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Here is a subset.

**Madhavan Mukund:** So maybe let us take one-third of them, 10 of them. Just take every third one.

**G Venkatesh:** Just take one-third, I want to keep it in order. Okay, this should be random enough.

**Madhavan Mukund:** hopefully. So, we should get 10 cards out of this to work with.

**G Venkatesh:** And if it worked for 10 we are saying that it might work for all, we do not know.

**Madhavan Mukund:** Okay, let us say...

**G Venkatesh:** because 12, because the months are 12.

**Madhavan Mukund:** Yeah, we will only see 10 months.

**G Venkatesh:** We should not take 10 months as 12. Let us take Jan to March, 4 only in a year.

**Madhavan Mukund:** so first...

**G Venkatesh:** Quarters..

**Madhavan Mukund:** Quarters okay. So, we have Jan to March. Then we have April to June, we have July to September then finally we will have October to December

**G Venkatesh:** So, that makes it 4, that is fairer.

**Madhavan Mukund:** Okay, so like grades are 4, these are also 4

**G Venkatesh:** Yeah.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Now, let us see what happens right we have taken 10 random...

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So, let us assume to be random. So, this is July

**Madhavan Mukund:** So, Priya comes here.

**G Venkatesh:** Priya goes there. Then this is February.

**Madhavan Mukund:** So, Surya goes in the first one.

**G Venkatesh:** Then January so Kavya is also there...

**Madhavan Mukund:** Kavya is also there.

**G Venkatesh:** Clash.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Clash, so will discriminate later.

**Madhavan Mukund:** Will discriminate later, we will keep on hold.

**G Venkatesh:** Bhuvanesh is November.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Monica is March.

**Madhavan Mukund:** So, she is also on clash.

**G Venkatesh:** Clash, JK is July,...

**Madhavan Mukund:** Another clash.

**G Venkatesh:** Ritika is November.

**Madhavan Mukund:** Another clash.

**G Venkatesh:** Sameer is March. Not coming other, April to June is not getting anybody, Geeta is May okay find a good one...

**Madhavan Mukund:** Okay.

**G Venkatesh:** And Rahul is April.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Okay.

**Madhavan Mukund:** Now, we have got a...

**G Venkatesh:** We got a...

**Madhavan Mukund:** Now we want.

**G Venkatesh:** We want a grade, then we look at gender.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Right.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** So, what do you have to do, now? You will see whether Surya, Kavya, Monica, and Sameer....

**Madhavan Mukund:** Yeah.

**G Venkatesh:** We take the Surya Kavya, Monica and...

**Madhavan Mukund:** Sameer.

**G Venkatesh:** Sameer, these are the four, okay, now we want to see whether these four can be...

**Madhavan Mukund:** Separated on grades.

**G Venkatesh:** Separated on grade.

**Madhavan Mukund:** Okay, so if it is Surya is 189 is a D.

**G Venkatesh:** Okay.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Kavya is 204.

**Madhavan Mukund:** Let us see.

**G Venkatesh:** Alright, so far so good, very nice Monica is 221.

**Madhavan Mukund:** Also see.

**G Venkatesh:** And they also have to behave with gender.

**Madhavan Mukund:** Yeah, And Samir is 250 which is a B.

**G Venkatesh:** So, gender does not work if you take months as if you do not take distinguish it by month.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** But Kavya and Monica if you take the month then January and March it will be a separator.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** You know you can take month after this after you do it in quarter.

**Madhavan Mukund:** The month you have done, no we want take town or month...

**G Venkatesh:** Town you can take, yeah town you can take Chennai and Bangalore.

**Madhavan Mukund:** so here town works but not gender, so, earlier gender was good for that but here town better.

**G Venkatesh:** Town better.

**Madhavan Mukund:** Okay.

**G Venkatesh:** See this tree, this tree says basically check whether or not they are on Jan to March equal to June, July to September, October to December.

**Madhavan Mukund:** Yeah, and now here...

**G Venkatesh:** Then it says take the grade, then it takes take the town.

**Madhavan Mukund:** Yeah, so here gender is the same for both Kavya and Monica. So instead, we will take town but then can you do that, can you sometimes ask different questions.

**G Venkatesh:** Depends on what path you take...

**Madhavan Mukund:** So, if they are born in January then we will ask grade and town, they are born in July we might ask.

**G Venkatesh:** Something else..

**Madhavan Mukund:** Something else

**G Venkatesh:** Why not? This is a tree you know where you come from right?

**Madhavan Mukund:** Why not?

**G Venkatesh:** So here we are town here and see if they separate it out ...

**Madhavan Mukund:** So, I will be cancelling it out, write town then I will get Bengaluru and kavya is Chennai, so...

**G Venkatesh:** So, this work so far for this set. Let us try the other.

**Madhavan Mukund:** So now we have look at Geeta and Rahul maybe they already have different weights because Rahul was a topper, I think...



**G Venkatesh:** I am sure Rahul was a topper, Geeta and Rahul?

**Madhavan Mukund:** Yeah.

**G Venkatesh:** 254 and 280.

**Madhavan Mukund:** They both got it.

**G Venkatesh:** so great. Did not work for them.

**Madhavan Mukund:** Yeah, so again, we have to ask one more question but again, we can do this...

**G Venkatesh:** town you can do...

**Madhavan Mukund:** So, both are here. But then if you do town as, then we will split Rahul and Geeta, so write town, so, what did not work for Srinidhi and Shashank, it will work for others, town

**G Venkatesh:** Then any other Clash, yeah, we have Priya and JK.

**Madhavan Mukund:** Priya and JK.

**G Venkatesh:** Priya and JK, alright, Priya and JK.

**Madhavan Mukund:** So, they have different grade.

**G Venkatesh:** So, they have different grades.

**Madhavan Mukund:** So, 187 so Priya has a D. So, if I go by grade, so, Priya has a D and JK has B so that is not a problem we already separate it out.

**G Venkatesh:** Okay.

**Madhavan Mukund:** Finally, we have Bhuvanesh and Ritika.

**G Venkatesh:** Bhuvanesh and Ritika.

**Madhavan Mukund:** And again, I think they have different grades because B and C. So, what was maybe I got Priya's grade wrong, please check Priya's 181 it should be D.

**G Venkatesh:** Yeah.

**Madhavan Mukund:** D and B and here it is C and B? So, this is Bhuvanesh and this is Ritika, so in two cases we were able to actually do it with two questions. And....

**G Venkatesh:** So, what is saying? We are saying that if you see if you see January to March, Date of Birth lying January to March and grade is D, then the person is Surya are trying to define we do not know we will see how to feel this from the sample.

**Madhavan Mukund:** Yes.

**G Venkatesh:** If the date of birth January to March and the grade is C.

**Madhavan Mukund:** Then it is Kavya.

**G Venkatesh:** And if the town.

**Madhavan Mukund:** Is Bangalore.

**G Venkatesh:** Bangalore.

**Madhavan Mukund:** Then it is Kavya.

**G Venkatesh:** Then it is Kavya. If it is C then it is Monika this is how we got so far. Let us just take couple of more cards from this set and see whether this formula this algorithm works, then algorithm.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** without doing anything,...

**Madhavan Mukund:** Yeah, correct.

**G Venkatesh:** We had to do something here some of you learnt it. this is nice. Actually, this is better. I like this.

**Madhavan Mukund:** So, let us see how it works?

**G Venkatesh:** Let us see how it works. Now we take this Rida that follows our algorithm. So may,...

**Madhavan Mukund:** We are in this one.

**G Venkatesh:** Okay. Then we are supposed to look at grade.

**Madhavan Mukund:** Yeah. We have noticed we have only seen an A grade.

**G Venkatesh:** So, we can put a new grade.

**Madhavan Mukund:** We can produce grid right so the grade is D, then this is Rida.

**G Venkatesh:** So, this goes for Rida no clash. Let me just take another random one from somewhere okay. Tauseef.

**Madhavan Mukund:** Okay.

**G Venkatesh:** 216.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Born in December, 216 is..

**Madhavan Mukund:** Grade is a C so same as Bhuvanesh.

**G Venkatesh:** So, we are supposed to use by the rules we do not, there is no rule for this..

**Madhavan Mukund:** No rule yet. So, we have to decide.

**G Venkatesh:** We will have to use the same, town.

**Madhavan Mukund:** It is a town. The town is Trichy. So, we have to go back and see what was Bhuvanesh's town...

**G Venkatesh:** Bhuvanesh is Erode.

**Madhavan Mukund:** Thank fully.

**G Venkatesh:** Okay.

**Madhavan Mukund:** so town we asked here then go to Erode. This is Trichy then we are separated on these two.

**G Venkatesh:** let us take one more random Arshad. December. here. Grade 210....

**Madhavan Mukund:** 210 is C.

**G Venkatesh:** Clashed. Chennai.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** Luckily.

**Madhavan Mukund:** So now.. So far, so good. So, we have these three questions we seem to work.

**G Venkatesh:** Try a couple more, Nisha, September,...

**Madhavan Mukund:** September. Grade is A.

**G Venkatesh:** B.

**Madhavan Mukund:** The same as JK. So now we have to figure out.

**G Venkatesh:** JK's.

**Madhavan Mukund:** JK and Nisha are separated by Chennai and Madurai. So again, we are survived.

**G Venkatesh:** Last one.

**Madhavan Mukund:** Yeah.

**G Venkatesh:** I think town again has too many possibilities.

**Madhavan Mukund:** We had state.

**G Venkatesh:** if had state then all these erode.

**Madhavan Mukund:** Okay.

**G Venkatesh:** Trichy and Chennai and all slide.

**Madhavan Mukund:** So, August,

**G Venkatesh:** August.

**Madhavan Mukund:** D, Priya .

**G Venkatesh:** Priya.

**Madhavan Mukund:** Vetrivel is here.

**G Venkatesh:** Priya is Nagercoil .

**Madhavan Mukund:** so, again, we survived.

**G Venkatesh:** Because again, because we have town is a huge discriminator.

**Madhavan Mukund:** Exactly.

**G Venkatesh:** It is got too much information. But this so this shows something that we are, I mean, this is demonstrating something to us? What does it say? It is saying basically that we can look at some examples.

**Madhavan Mukund:** And then build up this kind of...

**G Venkatesh:** Decision tree.

**Madhavan Mukund:** Decision tree, which is like an if-then-else in our notation of pseudocode. So, we can come up with some rules in the form of, if it is this, then it is this and this, then it is something otherwise, if it is this, and this and this, some other combination, it is this thing, and so on.

**Madhavan Mukund:** And this is this, some other combination it is this thing and so on.

**G Venkatesh:** Go on like that. Try it for a few samples. If it is working for those samples, make it work for the samples, then keep adding one element at a time. Yes. If it works for a sufficient number of such examples, you just take it as a working thing. Of course, it may be wrong, because it may wrongly identify somebody. Even this, they go through the full thing.

**Madhavan Mukund:** There may be clashes, which do not get resolved with these three questions. We say quite possible.

**G Venkatesh:** So in that case, it is basically identifying the person wrongly, maybe wrongly identifying or something. You may if, for example, here when we had Kavya, and Monica, and the each Monica, we may say it is Kavya, actually, but in cannot out to be Monica. So, you may wrongly identify Monica as Kavya. So, as long as the number of these false identifications is kept very, very small, reasonably small.

One or false identifications, alright, fingerprint recognition also is not 100 percent fool proof, not correct. So, that person will complain, and they will come to me. And then you will identify people by face or something, like mistakes.

**Madhavan Mukund:** Especially in that, so here, it is kind of because we have a very small set and we can easily check if two people are the same or different. But, face is very eminently possible or fingerprint that based on this, somebody could be wrongly classified as the person.

**G Venkatesh:** So, if I asked you this, we have these cards, we saw that we could use gender, we could use town, town has too much information, I think we can say, State something like that. We can say your birth month's information, but we could just divided into quarters. Our first half and second half, gender male and female, grade ABCD. Say state some lots of district maybe.

**Madhavan Mukund:** Because many of these are from the same state I think one many are from Tamil Nadu.

**G Venkatesh:** So I think this is really that we will not use that data, we can say great formats, correct. So, suppose I take like this different properties of the student? The minimum information because to check it must be easy. So, carrying that information is very easy to do, then what is the minimum number of such variables that I need to check before I can find out the person uniquely? How do I? What is their problem? They are sitting there?

**Madhavan Mukund:** So, you want to make this diagram, have as few levels as few levels as possible. You do not want to ask you want a smaller tree

**G Venkatesh:** You do not want to have too many branches of the tree becomes very big.

**Madhavan Mukund:** That is why we cut down this 12 into 4.

**G Venkatesh:** 4, so few branches and lower depth. So if I have 4 branches, each node or two, if it is true that if it has depth of 3, that becomes 8. 4 means I may get 4 into 4 into 4, I get 64. That is also too many. But I fall into 2 8 into 2, 16. That is. Something like that. Cannot be 16. Because we have more than 16 cards 30 to 30 cards, it has to be at least 32. But, can I do it with 32? Which, which variables were I choose in such a way that by checking the variables in sequence, rather than a tree form, I can uniquely identify the person.

**Madhavan Mukund:** So, when you ask a question, like we saw here, it breaks it up into some groups.

**G Venkatesh:** it breaks it up into groups.

**Madhavan Mukund:** So, maybe you can look at each question and kind of look at how the groups or the groups are getting separated? Well, that is a good question. If many people are lying in the same group, then you are not asking you are not getting much information from asking that question. So, you can ask for each question, which is giving us more information about the separation, separation? And then we asked that question, go to the next step again, ask him all the questions, which are remaining, of course, now we have asked so when, when we come here, we know everybody in this has got the same..?

**G Venkatesh:** We did. Actually, we looked at the cards and figured it out. But you could do it systematically, presumably.



**Madhavan Mukund:** You can say what all questions are still available to me. So, we can decide on some questions we are allowed to ask. And among those we say which questions are still remaining? Because one question we have already asked, no point asking it again, because everybody's in the same bracket, same bracket. But now, if we ignored the birth month, then we can ask whether it is a town is a good question, or the grade is a good question, and so on.

**G Venkatesh:** So, you are saying basically what we do is, but we are doing it for a sample not only for a sample, so you take a sample. So, you have.

**Madhavan Mukund:** There is a sample and then on the sample, you might ask yourself, if my first question is

(Refer Slide Time: 34:48)



**Madhavan Mukund:** if I first start by asking say month, then I will get some, so we have this 4 quarters. So, I will get some 4 groups. So, I will get some group 1, group 2, group 3 and group 4. So, that will give me some split. If it is a good split, it might be equal, but it might be slightly unequal for this case it became 4 to 2. Now I can have another 4 way split like grade.

**G Venkatesh:** As the same sample?

**Madhavan Mukund:** For the same sample, and I will get some, some different groups. The same exact same cards and then we can ask whether this, this distribution is better than the distribution. We can try that?



**G Venkatesh:** I will, I will take 123456 cards is good. They were 10. I do not know.

**Madhavan Mukund:** 10 is I think better.

**G Venkatesh:** Let us take 10, here are 10 cards.

**Madhavan Mukund:** So this time, we will just count we do not know. We do not care how many. I mean, who they are. We would not write down who they are. We just want to know, which is giving us a more even more even split.

**G Venkatesh:** So, grades. What is this? So this is Jan to...

**Madhavan Mukund:** So, this is July. So, this is the third group. So, I put a tick here. So, I put a mark here like we did the first class. I know and we also deal with this idea. So, grade is a D, so this is one person has got a D and was born there. Then this was again D, D. And in May so in the second group, August is a third group. Again a D so I think you picked up a lot of bottom up people in this place.

**G Venkatesh:** This is not working.

**Madhavan Mukund:** Again a D it was from December, 209 is a C from March. So, now we are getting a better distribution here. 219 is again, a C only I think? Think we had our grades? I think up to 227 was the C 200 to 226 is C C and October the last one. 24 again a C and September the last one again no the third one to

**G Venkatesh:** 24, exactly

**Madhavan Mukund:** 24 and in December. One more now this is a B. And in May. Finally, we have 244, which is still a B and in July

**G Venkatesh:** So, the sample is chosen? I mean, we did not choose it currently. But regardless, it says that just giving

**Madhavan Mukund:** Month is giving a little bit of better distribution,

**G Venkatesh:** Slightly a better distribution than this one. Because here for one group is not being represented at all. So, you should go with this. Ideally,...

**Madhavan Mukund:** Ideally, we would ask based on the size, you say that month is a better question to ask them. Great. Now we can also say, first question.

**G Venkatesh:** And then you could do the same thing at the next level.

**Madhavan Mukund:** Now I can say, of course. Now here, the sample is small. But how does?

**G Venkatesh:** Again, you can pick a sample and do the question from beginning.

**Madhavan Mukund:** Yes, you can do that also.

**G Venkatesh:** But we will have to pick people who are in this group.

**Madhavan Mukund:** To ask a question. So, you have to take all the people who are born between say, what is this July and September? And then ask what is the next question for this time that we can take out everybody who was born between July and September? Hopefully there are a few. This is July. September, August, July, July, September, October, July, July, February, March, January, May, April, September, September, September, May, we have so many. So, now we have all these people. The full set if we had applied month we would come down here to...

**G Venkatesh:** All of them are accomplished.

**Madhavan Mukund:** And now we want to ask the...

**G Venkatesh:** Which question to ask. So, the choices are great.

**Madhavan Mukund:** And so one choice is great. Other choices town and city but that we already decided is too specific. We have gender of course, which is only a two way split.

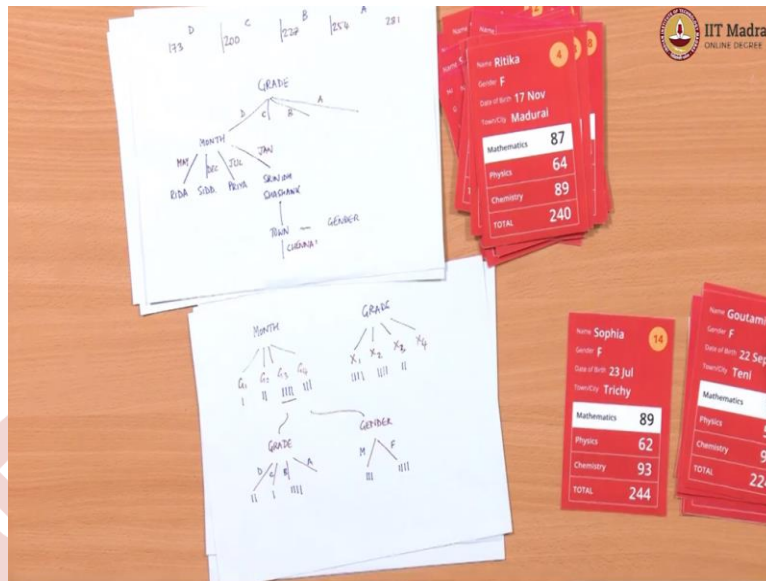
**G Venkatesh:** Two way split will clash anyway.

**Madhavan Mukund:** So, you can see.

**G Venkatesh:** I mean, we will try gender, gender might split it into two equal groups. That might actually be good.

**Madhavan Mukund:** So, we want to know whether it is going to be an unequal split or an equal split. That is what you want to know. So, so now the grade, 252 is a grade thing to do.

(Refer Slide Time: 40:08)



**Madhavan Mukund:** 252 is a B, I think this is our grades. So, 252 is a B, so I have a person who is got a B and was a male, let me write M, F, this is D, C, B and A.

**G Venkatesh:** 227,

**Madhavan Mukund:** 227 is also a B is also a male, 240 is also a B, this female 181 is a D and this female 196 is a D and is male. 224 is a C and is female. And finally, 244 is a B and is a female,

**G Venkatesh:** looks like gender is good.

**Madhavan Mukund:** So, gender gives us. And this is again, giving us a more unequal split.

**G Venkatesh:** So, this way we can do so the idea basically would be to try and pick if you are asking good ask a question. You do not ask that question this will distribute for the answers. If you have three answers, multiple choice like 4 choices directly,

**Madhavan Mukund:** We will try to put meet them into coal buckets,

**G Venkatesh:** it will distribute that equally.

**Madhavan Mukund:** So, that could be one strategy that we follow.

**G Venkatesh:** So, whichever out of the questions which one tries to distribute the most,

**Madhavan Mukund:** Evenly?

**G Venkatesh:** Evenly?

**Madhavan Mukund:** Would be the best questions.

**G Venkatesh:** Question to ask at that stage. Then you keep asking questions. And hopefully you ask as few questions as is required need to ask? And you end up with answer. So, this would be a Bottom up Algorithm.

**Madhavan Mukund:** Because we are deriving this. So, this tree finally, when we draw the actual tree like we did there, we have essentially a program, we have some kind of procedure, which tells us how to classify student. And we got the procedure by looking at the data, we, we did not decide in advance that we should ask grade first, so we should ask month first. So, in that way, it is not something that we analysed and said, this looks like a better one that because if the sample is different, we will get a different answer.

So if the data changes, then this tree will look different. So, the so we have a strategy to build a tree, we do not have it, but the tree itself comes from the data, the tree, we know that finally we are going to get something that looks like this, but what are the exact questions that we asked at each stage, those are determined by the data. So, that part is Bottom Up.

**G Venkatesh:** So, they say learning, this is learning algorithm?

**Madhavan Mukund:** So, this is this is how it works. This is more or less how it works? So we have a so we have a broad idea that we are going to be able to discriminate these students based on asking few questions about the data that is on the card now, which questions to ask is learned from the which are the minimal questions. So, you think about a doctor, no doctor is doing the same thing.

So, there will be you will come and you will have a lot of symptoms, you will say I have my head hurts my stomach is painting, I have a headache, my right shoulder is stiff. And then you will take a test and he will come with some blood test report with some other test report and all that. So, there is a whole mess of data, data there, what the patient is saying plus what the patient's reports are saying, but the doctor has to be smart enough to understand that this these 4 things together, identify some user some, some disease or some condition. And the rest is all just some, maybe this person is generally, suffering from some, some of

**G Venkatesh:** You also got a screen or something.

**Madhavan Mukund:** So, all these things are not, not relevant to this.

**G Venkatesh:** They are not relevant to this particular.

**Madhavan Mukund:** So, that goal of so, so there again, the doctor is using the kind of they are the data,

**G Venkatesh:** Ask some questions also?

**Madhavan Mukund:** Ask some questions, but the for the doctor this data is coming from the medical experience that the doctor has, the doctor has seen so many patients as they learn so many cases,...

**G Venkatesh:** They know that the person with fever, these are the possible diseases. So, these may be some additional symptoms for them.

**Madhavan Mukund:** So really, we are doing a kind of diagnosis the diagnosis like that. So this is like diagnosis. So, we are trying to say but the question is that like an expert doctor, which questions to ask. So, bad doctor will just say go and take 100 tests and come back. The good doctor will try to identify which are the relevant tests and I want to rule out this. So, you take this test or I want to confirm this so you take this test so the doctor will

