Course Name:Business Intelligence and Analytics
Professor Name:Prof. Saji.K.Mathew
Department Name:Department of Management Studies
Institute Name: Indian Institute of Technology Madras
Week:01
Lecture:04

COURSE OVERVIEW | BI&A

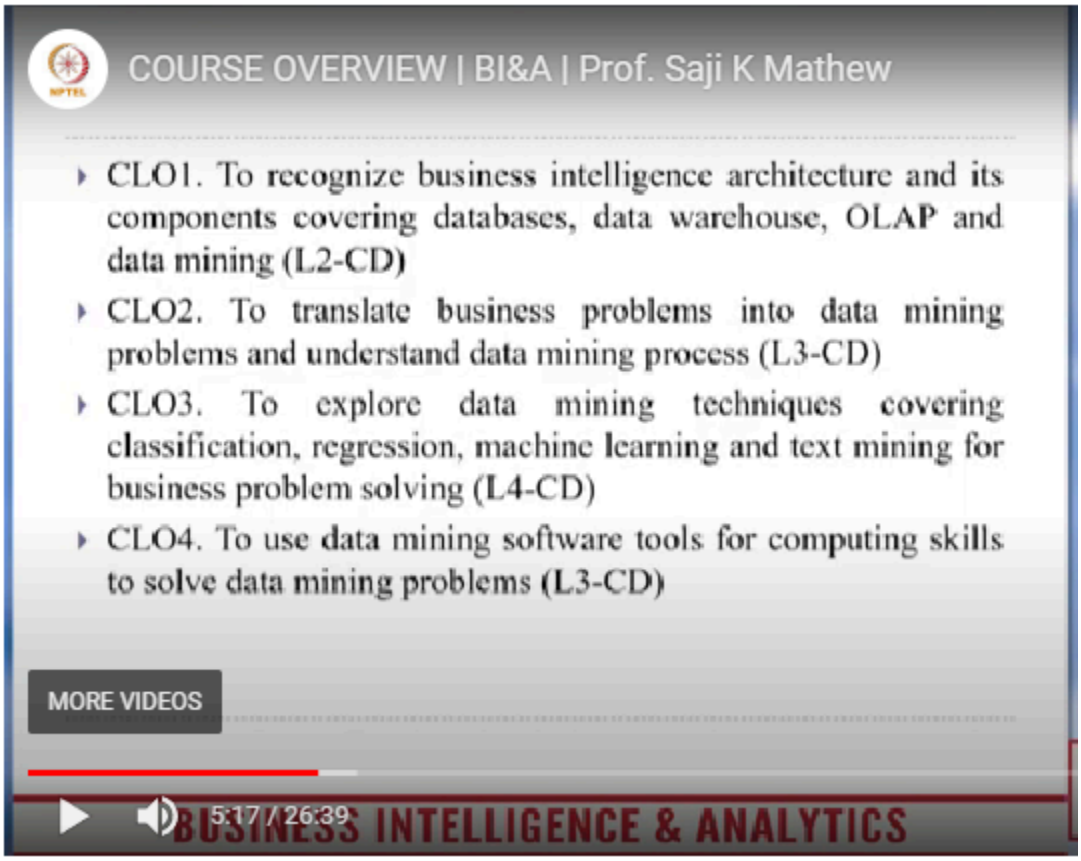
Hello, and welcome back to the session on Business Intelligence as a part of the Business Intelligence and Analytics course. So during this initial sessions we are getting introduced to some of the fundamental concepts related to Business Intelligence and Analytics. So we have seen how BI&A evolved from information systems as a separate layer or a separate entity which provides certain solutions to industry and not only industry but different domains of human life and enterprise. So our purpose in this session is to make a clear understanding, develop a clear understanding about various components of Business Intelligence systems, Business Intelligence and Analytics systems. Okay so let us treat them together. And also in doing so develop clear clarity of definitions of various concepts and understand where each concept or each component from an architectural point of view where each component fits in. That would be our effort today.

Another aspect of understanding BI&A from a fundamental level is to understand the curious connection between business problem and Business Intelligence problem or business problem and analytics problem. How problem which is purely business related can be converted into an analytics problem. And then how you pursue that problem as an analytics problem is in the domain of analytics. But bridging this business to analytics is a major effort and that part is the thought process.

So you think about a problem in detail and try to discover the analytics aspect or analytics lens, you know you apply an analytic lens on the problem, you try to look at the problem through a different perspective. And then discover solution using that perspective. So therefore we are not claiming that analytics is not the only way to look at a problem. There could be other ways also of looking at problems through different frameworks. So through different disciplines and their efforts.

So database decision making is the core theme of this course. So we will take that perspective in addressing problems faced by business and management. So initially let me actually give you a walkthrough of the course which I did not do previously. I only introduced you to what is BI, what is analytics and how these concepts evolve, what are the current drivers for making it a discipline in itself today or a sub discipline in itself

today. So let me actually, before we go further into the details of the session, develop better understanding of what will be covered more specifically in this course or what are the objectives of the course.



So the objectives or learning objectives of this course are fourfold or there are four learning objectives. The first objective CL01 course learning objective 1 is to recognize business intelligence architecture and its components covering databases, data warehouse, OLAP and data mining. So this is according to Blooms taxonomy, it is level 2 and it is cognitive domain that is L2CD to define the objective more clearly as to where it operates or where it is situated. So essentially the objective is to give a clear picture of the technical architecture or the technical architecture which is the plan and the infrastructure elements that are involved in creating data based solutions or data driven solutions to business. So if a business is serious about BI &A, then it needs to have certain assets, certain resources within the organization that includes a database and it includes a data warehouse , where you warehouse data.

So the understanding here is data is the raw material as we discussed previously, data is the raw material and therefore an organization thriving on BI&A should have proper means to store its data for the purpose of analytics, not just for transactions for the purpose of analytics. So therefore the architecture or the plan of the organization for

BI&A becomes very important. So we will be spending a session on data management, we will have another session on OLAP and data warehouse. So and subsequently we will move to the data mining or analytics. In this course, I will be using this terms almost synonymously data mining and analytics, essentially both will include use of advanced techniques to derive insights from data, to model and derive insights from data so that those insights could be applied for solving business problems.

So it is in that perspective we look at data mining and analytics. So essentially the first objective aims to provide a clear picture of the technical components or the technology pillar of analytics, as I outlined in the previous session. So the next session, sorry the next objective is to translate business problems into data mining problems and in turn understand the process of analytics or data mining. So this is a very very important objective and it is very hard to teach this through lectures. This is something one has to learn through practice as well.

I can underline principles of how to convert a business problem into an analytics problem but that may have certain context. When you move from, say healthcare industry to a banking industry, you will find the domain very different and therefore the type of problems are different and the way you have to think through the problems and the availability of data, external and internal all that could be different in different domains. So translating business problem into a mining problem or an analytics problem is a thought process. But what essentially is involved in that thought process is what you will learn. So I will be spending a session, in fact more than a session to discuss a case and see how a business problem got converted into analytics problem.

That is a case based illustration and then we will also learn certain fundamental lessons or a fundamental principles in data mining or analytics process. So there is a process as is shown in CL02. So what is that process of moving from a business domain to a analytics domain to solve business problems. So we will learn it as a stepwise process that will bring more clarity. And thirdly how I drive this objective is through a number of other problems we will be discussing in the course.

So which will be solved using data. So I will be demonstrating solutions for business problems through data and data mining or analytics. So through this lessons, you must be able to get a sense of what is this thought process in understanding and translating problems from business to analytics. That is the second objective.

And third objective CL03 is to explore data mining techniques covering classification, regression, machine learning and text mining for business problem solving.
Now as a part of this course, you will be learning a set of specific methods or techniques or algorithms. So whatever you want to call them but specific techniques that are useful

in solving certain types of, certain category of problems. This includes classification. Classification is a general problem and for classification there are a number of algorithms or techniques available today both from statistical community as well as from algorithmic or computer science community. So we will be using or I will be explaining to you both.

So statistics meaning, you know the regression sort of techniques which were developed in statistics, whereas a decision tree or a neural network etc. were developed by the algorithmic community or CS as we call them. So we will combine both in solving problems in analytics. So you will learn decision tree as a specific techniques in this course. You will also learn to apply regression analysis by, machine learning is a more generic term today.

Your textbook one of the recommended textbook is titled Statistical Learning. So these titles of statistical learning, machine learning etc. are very related in the sense algorithms or machines are able to learn like human beings. Learning is not restricted to humans but machines can also be trained. So in that sense today if you look at general textbooks on machine learning, everything is put under machine learning including say regression or logistics regression or classification techniques, neural networks.

And combined today you call them AI, ML etc. So machine learning is basically involves a training algorithms to learn patterns in the data and then make them interpret or predict for using a different set of data. So that is essentially what is involved in machine learning. So I have mixed up those terms in CL03 but you would be able to make sense when we go through certain specific techniques and you know certain problems that can be addressed using those techniques. So I would say the emphasis of, the emphasis of this course is on problems.

What problems you solve? So analytics as a domain or a sub domain which addresses problems of the world using data and data analytic techniques. So therefore from that perspective, analytics is a problem solving enterprise or a problem solving domain. And we take that perspective. So therefore problem comes first and then we look at different types of problems like classification is a problem that business faces. You need to segment a product, you need to segment customers, you need to segment geographies and so on.

So group things. So then for that purpose for or that class of problems or type of problems, what are the techniques that is available. So we look at, so we keep techniques at the back end and we keep business and management in the front end. So that is the philosophy of this course. Instead of turning it going to the back which is actually the view, the technology view or the computer science view or the technical

community works on developing best algorithms, testing them or improving them. But we are not attempting to do, improve algorithms in this course but we are attempting to use algorithms but in doing so we must be aware and quite aware of how algorithms work and what is the logic that is embedded into the steps of algorithm.
What is it trying to do? That clarity is very important in applying algorithms.

So I am not undermining the importance of algorithms in analytics but the point I want to stress is algorithms are used to solve problems. That is the perspective. So our effort is again not to design new algorithms as I said but to use algorithms effectively. To choose and use algorithms effectively.That is CLO 3.

And CLO 4 the fourth objective of the course is to use data mining software tools for computing skills to solve data mining or analytics problems. So what would that mean? To use data mining software tools for computing skills. So it emphasizes the need for learning certain software tools as part of this course. Keep in mind that analytics and business intelligence uses large volumes of data, implying this data do not originate from surveys or experiments where the size of the data is very small. You can actually use data analysis tools or the simple tools to analyze those data statistically.

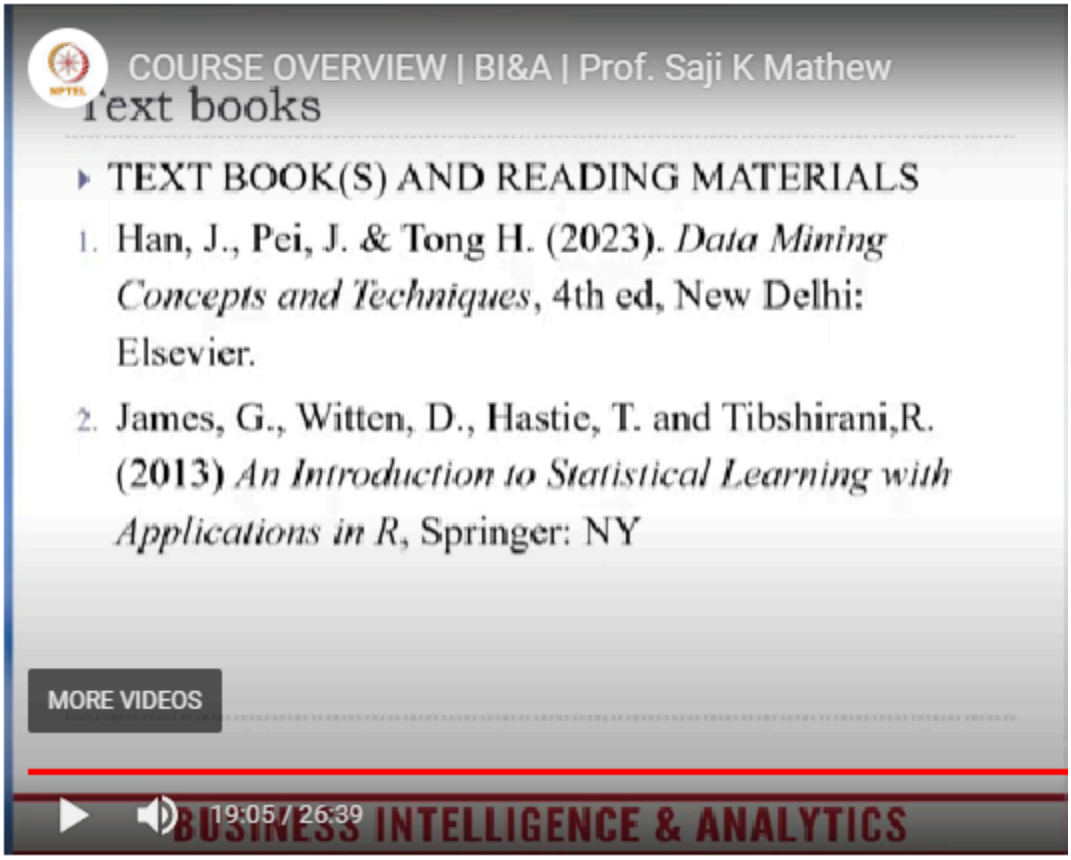
But when it comes to mining, the word itself imply that it is huge. It is like vast amount of unrefined raw material. And that is the sort of data that you will be dealing with when it comes to enterprise systems, enterprise databases or even external big data databases and so on. The data volumes are too large and therefore you need different kind of data mining tools to deal with such large volume of data. That is one aspect and therefore use of software is imperative because manually you cannot solve or analyze that data.

You need software and therefore this course emphasizes the need to learn software. Now let me again emphasize we will be using tools, software tools in this course but the training or the sort of tutorials for learning those software tools is not part of this course. So software learning is up to the participant. So but as part of this course, I will be demonstrating a lot of applications where we use software. I will show you which are the specific ones we will be using.

So let me summarize the course's four objectives right from learning the architecture of BI&A to understanding business problems and translating them into analytics problems and then applying algorithms to solve those problems and also learning software tools that will be useful in analyzing data because these are the objectives of the course.

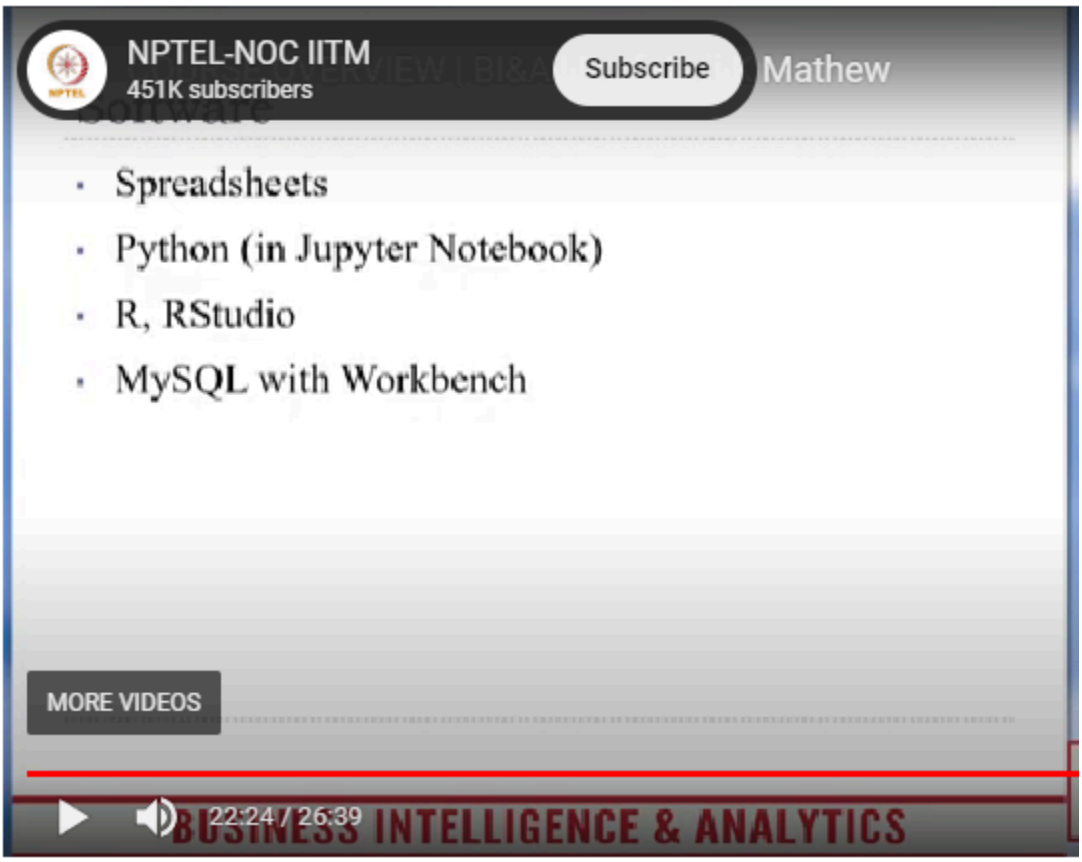
So let me also introduce the textbooks more clearly. There are as you see, there are two textbooks that will be used in this course and it is mandatory to have these textbooks if you are crediting this course because your assignments and your sort of questions that you get could be related to the textbooks. So it is not one textbook, I refer two books one

is Han, Pei and Tong 2023, Data Mining Concepts and Techniques, fourth edition, Indian edition of this textbook is available and it is economical and second book as I said previously also is James Witten, Hastie and Tibshirani and this book is very widely referred today for an introduction to analytics and a soft copy of this textbook is available online. If you search, the authors are very graceful in keeping the book soft copy online but it is up to the participant to go and get a copy or you can buy this book. Indian edition is not available to the best of my knowledge, so it may be costly.



Now I said we will be working with software tools in this course. Which are the ones that will be used in this course? Of course spreadsheets meaning you can use excel or a Google sheet but essentially we will use, particularly for demonstrating the data and how it is structured and sometimes you know, you can use spreadsheet itself to do basic analysis if the size of the data is limited and if you are using basic techniques like t-test or regression etc. you can use spreadsheets and we will be using Python and Python in Jupyter notebook. So the Anaconda distribution of Python is what will be used as part of this course but there is time for the specific use of the software. It will be towards the end and so is R and R studio particularly, you must have noticed we will be learning text mining as part of this course. There is one module on text mining techniques. So, and after introducing the technique or text mining as you know, as an area of analytics, we will be working on one problem, where I will use R studio and certain libraries will be used in R studio and this will be shown to you in advance as to where you can get access

to this software etc. will be available in your course announcement.



So therefore you can look at that and get the software installed in your machines and MySQL with Workbench will be one session where we will again demonstrate you the use of relational databases and how you can store data in a structured way in relational databases and also how you can query those databases. So SQL, structured query language will on certain table stored in a MySQL database. MySQL is the commercial, you know it is the database technology.So we would be demonstrating it to you and therefore MySQL is also a part of the software tools that will be used in this course. So all this will be demonstrated to you and you can in parallel learn to use this software based on your interest, of course MySQL is up to your interest.My purpose is to demonstrate how structured data are stored and how they can be used for sort of descriptive analytics using queries. So one feature of all the software I have listed is that they are all free software.You do not have to buy licenses of the software as was the case say, maybe 10 or 15 years ago . So you have to buy license for a MATLAB software which is quite expensive but very advanced in its features I agree, but today you can get the same features from software, programming software like Python, you know Jupyter notebook or R studio etc. These are all free open software tools which are used as part of this course to model data.

So that is about the use of software in this course. Now certain data sources will be used in this course. The first two items mentioned here are sort of proprietary and it is not shared with the participants but it will be demonstrated to the participants because they are sort of proprietary data and but the other data sources like the UCI machine learning repository, Yahoo finance, Twitter, ISL resources meaning the textbook itself, your textbook introduction to Statistical Learning has a open data repository, from which you can download certain data which the textbook uses and Kaggle as you are aware today is a platform for analytics and they also provide access to certain pilot data sets. So, these are all open data sources which can be used by the participants for, you know working on problems or practicing on certain analytics techniques etc. And you are encouraged to visit and explore these data sources and as I said, the first two are not open but others are. So please do make use of them and there could be more but these are the ones I am well aware of. So that is the summary of the course and you will be seeing more about the course, the sessions and other materials that will be used in the course etc. when the course is announced.

So, and at the moment I would stop there and now our aim today is to gain a clear understanding of the business intelligence architecture and also the process of translating business problems into analytics problems. So from that perspective, I would want to

show a case, a short case and we will read that case and analyze that case using a few questions and then subsequently we would develop our understanding about the thought process in translating problem from business to analytics and also the infrastructure need for problem solving using analytics. So, that is the starting point or that is where we begin and then we move forward. So, let me show you the case and spend some time in reading and understanding the case. Thank you.