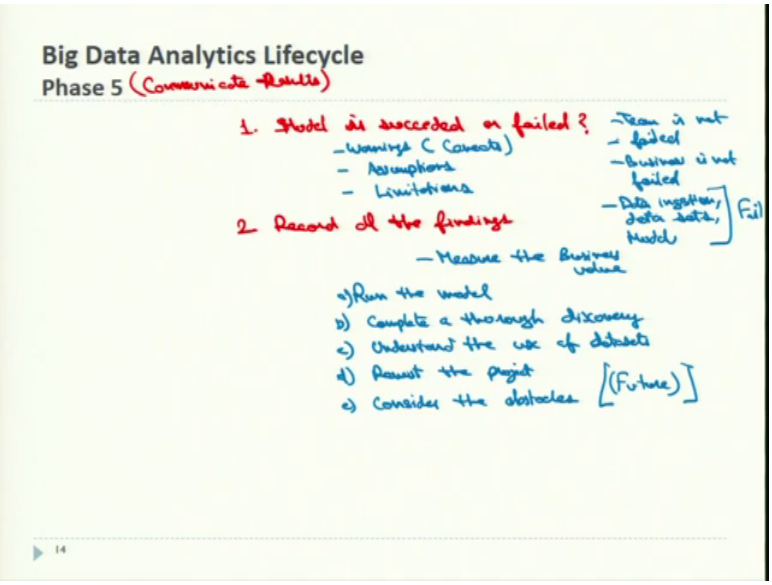
Computer Aided Decision Systems
Industrial Practices Using Big Analytics
Professor Deepu Philip
Department of Industrial and Management Engineering
Professor Amandeep Singh
Imagineering Laboratory
Indian Institute of Technology, Kanpur
Lecture - 21
Big Data Analytics Lifecycle - Phase 5 and 6

(Refer Slide Time: 00:13)



Welcome, the next point is phase 5, in which we communicate the results. Communicating results means the model is selected, the model is ready, the model is built, and we have run the training data and we have tested it for validation. Now, we are trying to communicate the results before we actually go for the production. So, communicating results means

1) We try to see whether the model would succeed or fail- This is the first thing that we try to see here in phase 5. Many times, people do not want to admit to failing it. Here it is to be very clearly mentioned that if the failure is there, that is the failure of the data and not the team. So, the team has to accept failure and it is to be with no bars, it is to be accepted by the upper management as well and a team is to be told that the datasets were not right or the data itself is not putting the right way, it is not that the team has failed, but the model has failed because of some computational errors or because of some data input errors. The team can continue working while selecting some other models. So, models succeeded or failed. So, the model fails, I would say

- Team has not failed.

- Business will not fail, only the data ingestion or datasets are not representative or the model itself, these have failed.

Now, with this, we need to compare the outcomes of the modeling to the criteria established for the success and failure. And, with this criteria itself we need to see whether the criteria which was there is the model representing the model failing according to that or not.

So, taking into account all the caveats, assumptions and limitations, that is, we try to take into account all the

- Warnings (Caveats)
- Assumptions
- Limitations

So, can we leverage something in this? If the model has failed, can we leverage the assumptions or the limitations? Can we make it a little more open? Or if the model has succeeded in the first step itself, can we make the assumptions more strong or can we add some more assumptions? Can we add some more limitations? So, these things are also taken here.

So, while communicating the results, these calls are taken by the people at the upper level, maybe the business intelligence analyst or maybe the program manager, because the results are not to be communicated. So, the program sponsor, who would now be communicated the results, would ask the questions. He can pick any nuances and try to ask queries in that. For that, the things are tested time and again, and we keep on changing them. So, the limitations to a very small extent, so that we can answer or we can be able to get the final output while having input from the program sponsor as well.

So, again, the team must be rigorous enough with the data to determine, whether it will prove or disprove the hypothesis which were outlined in phase one, sometimes a team has only done the superficial analysis which is not robust or good enough to accept or reject hypotheses. So, the analysis is superficial. So, that is why the wrong entities are selected. That means it becomes an alpha or type one error once again or another. In the other times, the teams do very stringent analysis. It goes for a very robust model which tries to reject the variables which were otherwise, okay to it again because a type two or beta error the both of them should also be avoided. In this assessment, it is to determine whether the results are statistically significant or valid. And, we have to identify the aspects of the results that stand

out or we provide the selling findings. If the results are not valid, think about adjustments, which are just set in assumptions and limitations and which can be made to refine or iterate the model to make it valid.

Now, the best practice in this phase is to

2) Record all the findings- Let us select three most significant ones as we do the regression, that is in general and this can be shared with the stakeholders. We measure the business value, record the findings, then we

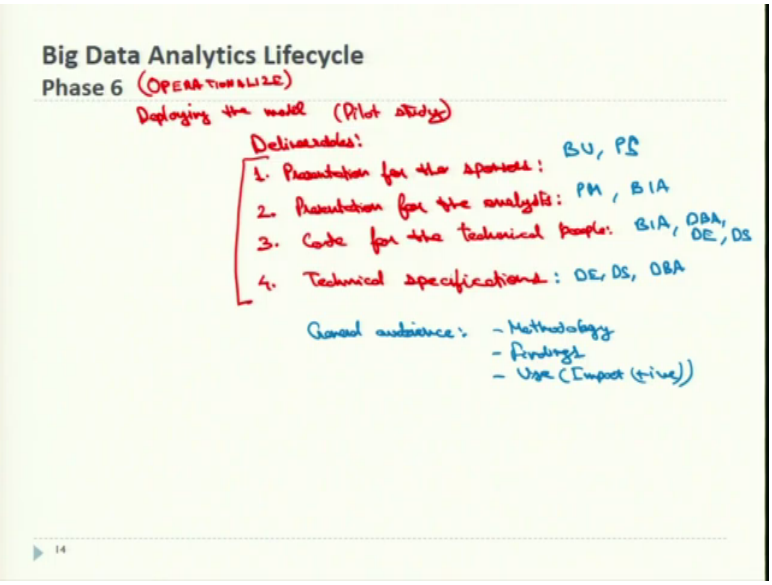
- Measure the business value here.

And depending upon what emerges as a result of the model, the team may need to spend in quantifying the business impact of the results to help to prepare the presentation or demonstrate the value of the findings. So, in this case, the team has to now run the model.

a) Run the model.
b) Complete a thorough discovery.
c) Understand the use of the datasets,
d) Revisit the project once again
e) Consider the obstacles.

So, we are in the project which can be improved in future. Revisit the project and consider the obstacles, which will come in future though we are communicating the results. Sometimes a model could be accepted in this form itself, but in the future, we are going to make small changes, so that can be taken into account here. So, the team will have documented the key findings and major insights derived from this analysis and we try to communicate that.

(Refer Slide Time: 07:41)



Next comes phase 6, that is the last phase, where we operationalize the analytics model. Operationalize means the team communicates the benefits of the projects or the model in a broader way and sets up a pilot project to deploy the work in a controlled way. So, we are deploying the model.

So, first we discover the business domain, the objective statements, the hypotheses, then we prepare the data. For that, we spent much time in the preparation of the data, then we try to plan for the model in the phase 3, then we selected the model on what was planned still if those were small changes, those were made, they were communicated in the phase 5 to the upper level of hierarchy, the sponsor or to the program manager.

So, then once the inputs are taken also from them, we try to now operationalize the model and we will deploy the model but this deployment of the model is also for a small set of the data that is a pilot study. This is done in a controlled way. The majorly in the operationalizing the model there are certain deliverables.

1) It is a presentation for the sponsors- Which means a high-level takeaway for the executive level stakeholders, a few key messages to aid the decision-making process to be focused on the very clean visuals like we present only pie charts, histograms, line diagrams. Very clean visuals are there, to which they could understand without even having much knowledge on these statistics or so. So, these presentations are prepared. All of these deliverables, which are there, do have the visualizations as well.
2) Presentations for the analyst- Because they could understand the more technical part of it, so, we try to describe the business process changes or report the changes. The

data scientists would like to also detail the comfort with technical graphs, where we can have a boxplot and we can have fan charts. We can have bubble plots or so, and little statistical techniques more sophisticated than which are presented to the sponsors are provided here.

3) Code for the technical people- The code is given for the technical people, which means the actual code, which is written for the program to run the pilot study, is developed by the data scientist. It is vetted by the data engineer and presented to the upper level and finally it is approved, it is tested, it is run, it is not prepared for the technical people to run it and put or ingest the actual data which is to deploy the model.
4) Technical specifications of the implementation of the code- When I say technical specifications, here an algorithm is written that shows how a code would be actually deployed in a realistic situation. It is a kind of an operational manual for our machine. When I say machine, it is a machine learning algorithm for which we operate? Where do we input the data? Which are the data points, which could be altered? Which are the text variables here? Which are the variables that could be varied as and when required?

So, all these techniques specific cases, that would help to these four documents, this is a document. Only technical specifications to these four deliverables are there in the operationalizing phase. Here, each person, business user, the project sponsor, project manager, business intelligence analyst, data engineer or database administrator, data scientist, all of them have their own roles to play here.

a) Business users here help in presentations for sponsors as well, it determines the benefits and implications of the fundings of the business and helps to prepare the presentation for the sponsors. The Project Sponsor itself, as the cost, is related to the return on investment, how the business will impact the final project.
b) The Project Manager needs to determine if the project was completed on time or not. So, in the presentation for the analyst, the Project Manager plays a major role here, when the project sponsor tries to have a look at the report, the project manager sees the timeliness of the project or the milestones are all met or not, is there any change in the milestone which was pre laid. Then comes the business intelligence analyst. It sees whether the reports and the dashboards he has managed, has impacted or do we

need to change them or not. So, here is the technical code, which is written for you. I will put business intelligence analysts to work in point 2 and 3.

c) And also, in the code he tries to see that what codes this person has given here or supplied here are actually used or not and data engineers and database administrators need to share their code. So, database administrator and data engineer share the code, and data scientist is also part of it.

d) So, both, data scientist and data engineer prepare a specification report, which is further vetted by the database administrator as well sometimes and this is prepared accordingly. So, running the model, whether it fits smoothly into the production environment, this can be integrated into the latest business processes. Then creating a mechanism of performing ongoing monitoring of the model accuracy. If accuracy decreases and finding the ways to retain the model or there are certain design alerts, as I said warnings etcetera, are to be also put if the model is operating out of the bounds even. So, all these points are taken into consideration when we try to operationalize the model.

So, while presenting the use of the model to the audience, audience means the people who are not part of the data analyst team. So, more focus is to be given on the findings and the use of the model than what would be the impact on the business. What are the findings of this model? So, this is presented to these four deliverables which are laid down here are only for the team or only for the stakeholders, who were selected in the beginning itself.

Now, the general audience are given quantitative backgrounds, that is, what is methodology? What is finding? So, this audience will be more interested in the techniques especially if the team developed a new way of processing or analyzing the data that can be used in future or apply to similar problems. So, I would say,

- ❖ General audience: We will try to give them

i) Methodology

ii) Findings

iii) Use or final impact

In addition, imagery data or visualization is always possible in place of putting the bullet points. One can put images and try to say okay; these are the outputs that could come, these are the issues that could come, these are always presented. So, this is all operationalized models here.

So, with this, all the stages are covered. These were only the discussions about the phases. This will definitely make more sense. When I will try to come up with a case study, we will try to have a case study in manufacturing in detail. I will also talk about the design thinking, how design thinking itself is talking about the lifecycle phases, where we will also see the data visualization, and we will meet next week. We will be discussing more on that post on Computer Aided Decision Support Systems. Thank you.