

Career College, Bhopal

Lecture notes: Data Science Process

Define the problem: The first step in the data science process is to clearly describe the problem that needs to be solved. This means understanding the business context, figuring out who the stakeholders are, and figuring out the project's goals and objectives.

Collect and clean the data: The next step is to collect and clean the data that will be used in the analysis. This step involves getting the data from different places, making sure no data is missing or wrong, and getting the data ready for analysis.

Explore the Data: Once the data is ready, the next step is to look into it and learn more about how it is put together and what it is made of. This step involves making charts and graphs and using statistical analysis to find trends, patterns, and connections in the data.

Prepare the Data for Modeling: The next step, after looking at the data, is to get the data ready for modeling. In this step, you choose the important features, change the data into a format that can be used for modeling, and split the data into training and test sets.

Build the Model: Building the model is the next step. This step involves choosing the right statistical or machine learning algorithm, training the model with the training data, and tweaking the model's parameters to make it work best.

Evaluating the Model: In this step, you'll look at how well the model is working and compare it to the goals and objectives you set in the first step. This step helps figure out if the model is correct and reliable enough to be used.

Improve the model: If needed, based on the results of the evaluation step, the model can be fine-tuned and improved. In this step, you might try out different algorithms, change the parameters, or add new data.

Deploy the Model: If the model meets the performance criteria, it is put to use in the real world. This step is about putting the model into the systems and processes that the organization already has.

Maintain the Model: The last step in the data science process is to keep an eye on the model and make sure it's still working. This step involves keeping an eye on how well the model is working and making any updates or changes that are needed to make sure it keeps working well.

Summary: In conclusion, the data science process is a cycle with a number of steps, from defining the problem to keeping an eye on the model and making sure it works. Effective data science requires people from different fields to work together, have a deep understanding of the problem, and be able to turn data into insights that can be used to drive innovation and help people make good decisions.

Links for further reading:

<https://www.coursera.org/lecture/big-data-introduction/steps-in-the-data-science-process-Fonq2>

<https://www.youtube.com/watch?v=s4uF8UOJz9k>

<https://www.youtube.com/watch?v=Y7axWbf5haI>

Dr. Sanjeev Gour

Dept. of Computer Science