Scout (Mugdha) Sujyot

Week 2 paper

What are the top 5 ideas you retained from the material read in chapter 1?

1. Data science projects have many roles. In addition to the data scientist, there is the project sponsor, the client, the data architect, and operations. These roles represent business and client interests, as well as operational concerns.

2. The project sponsor represents the business interests and champions the project, while the client represents the end users' interests. The data scientist is responsible for taking all the steps to make the project succeed and the data architect is responsible for all the data and its storage. Finally, the operations role is responsible both for acquiring data and delivering the final results.

3. The stages of a data science project loop and overlap each other. The can be roughly defined as "defining the goal", "data collection and management", "modeling (or analysis)", "model evaluation and critique", "presentation and documentation", and "model deployment and maintenance".

4. For a new data science project, it is very important to set realistic expectations for all stakeholders. For instance, you should make sure that the resources you have are enough to meet the goals.

5. There are several important criteria when estimating if your data is good enough to meet their goals. It is important to understand how well a model should do and how well it can do. The null model is a lower bound on performance, and the bates rate is an upper bound on the performance.