Sean Cordrey

Sept 5, 2014

Roberts

Stat 3654

Weekly Paper 2

Chapter 1 of the reading is predominately concerned with a data science project and the people and procedures involved conducting a successful project. The roles include a sponsor, client, data scientist, data architect, and operator. The sponsor is the one who wants to obtain results from a data science project. Clients are end users who will use the results. Data scientists design the project, find sources of data and decide on which tools to use. Data architects maintain and store data. Operators put into action the solutions found through the project. A project has a number of steps which are not done necessarily in a certain order. Throughout one may need to go back a step to reevaluate before continuing on. Something I think is important is when setting goals, that the goals are not vague. Goals should try to be quantifiable so that solutions can be tested to determine success or failure. Also like any other project communication is key, and as a data scientist, it is important to remember that many may not have the same technical background so when communicating with sponsors and clients, information needs to be relayed in a way they can understand. Lastly, it is crucial that throughout the project goals are reconsidered for feasibility. After working with the data and resources, one has a better view on whether or not given the current level of data and resources, the set goals can be met. This check makes sure the current resources are sufficient or if more need to be sought out, and if more are needed, can they be obtained.