← Go Back to Capstone Project

Course Content

Milestone Submission

The milestone focuses on the initial exploration of the problem and the data, to come up with an initial understanding of some patterns in data, and the first set of results from the analysis so far including both, the insights from the data relevant to the problem as well as the results/outputs from various different modeling techniques explored.

The milestone submission should focus on the following two aspects. The first aspect is to explore data using various techniques learned in the program so far. It could be something as basic as analyzing individual variables through univariates or analyzing the relationship between the dependent and independent variables using bivariate or deep-diving into correlations between various data variables or more advanced unsupervised learning methods like principal component methods or clustering methods. The purpose is to arrive at key insights relevant to the problem

The second aspect is to have an informed view on what modeling techniques will work the best. To identify what techniques will work the best, it is imperative to try out different relevant approaches and compare their results to assess the relative performance. The assessment of various approaches should also have due consideration given to the implementation feasibility, i.e., a view on the potential challenges & risks associated with implementing a particular solution and the likely benefits or some measure of the potential outcome. The purpose is to have a strong understanding of, and a well-informed view of what the final solution will look like and why that solution makes sense.

. Problem Definition
. The context - Why is this problem important to solve?
. The objectives - What is the intended goal?
. The key questions - What are the key questions that need to be answered?
. The problem formulation - What is it that we are trying to solve using data science?
. Data Exploration
. Data Description - What is the background of this data? What does it contain?
. Observations & Insights - What are some key patterns in the data? What does it mean for the problem formulation? What are the data treatments or pre-processing performed?
. Building various models - This involves constructing multiple relevant models to solve the problem and evaluating their performance. The models can be fine-tuned to determine whether their performance can be improved.
. Comparison of various techniques and their relative performance - How do different techniques perform? Which one is performing relatively better? Is there scope to improve the performance further?
. Proposal for the final solution design - What model do you propose to be adopted? Why is this the best solution to adopt?

Review Parameters	Review Points
Problem Definition	4
Data Exploration	4
Building various models	4
Comparison of various techniques and their relative performance based on chosen Metric (Measure of success)	4
Proposal for the final solution design	4
Total	**20**

Previous Next