Assignment-1: 3 questions

Due date: 12/13/2020

11/25/2020

Project Title: The First Step To Data Science

1. Based on the videos and the reading material, how would you define a data scientist and data science? (3 marks)

Data science is a process of finding valuable insights, trends, and patterns that are hiding behind the large quantities of data and communicating these insights to stakeholders so that they can make strategic choice. Data Scientist is one who deals with Data Science.

Today vast quantity of structured and unstructured data are available from varied sources such as log files, email, social media, health data, sales data, sensor data, and so on. Data scientists use data analysis to reveal valuable insights from these data in order to help organizations understand their environments, existing issues, and hidden opportunity. Data Scientists help organizations resolve their specific problem and take data-driven decision.

In the data science process, Data Scientists perform data extraction, data cleaning and transformation, explore trends and patterns using different models and algorithms, and showcase the results using data visualization.

Thus by analyzing data and communicating new knowledge and actionable insights, Data scientist can lead the organizations to a new approach.

1. As discussed in the videos and the reading material, data science can be applied to problems across different industries. What industries are you passionate about and would like to pursue a data science career in? (1 mark)

I did my M.Tech in Electrical Engineering. My M.Tech thesis was related to Power Network Uncertainty Analysis. I wrote codes and performed statistical data analysis for my thesis. My interest lies in the power/energy industries and Tech/IT industries.

1. Based on the videos and the material, what are the ten main components of a report that would be delivered at the end of a data science project? (5 marks)

The final deliverable of a data science project is crucial to communicate new information and insights of the data analysis to stakeholders. Data scientist should produce a compelling report to promote their findings.

According to the Prof. Murtaza Haider, author of the book “Getting Started With Data Science”, the final deliverable, whether it is a brief report of few pages or a long detailed document of 100 pages or more, should follow a prescribed format including ten main components. These components are:

1. Cover page: The cover page should be informative. It should include the title of the report, name of the authors, their affiliations, and contacts, name of the institutional publisher (if any), and the date of the publication.
2. Table of contents: The table of contents (ToC) includes main headings, lists of the tables and figures, which give an idea of what lies ahead in the document.
3. Executive summary: The abstract or executive summary is the most powerful section where author can explain and argue the main points of the subject matter.
4. Introductory section: The introductory section gives an overview of the subject matter. It may include literature review to highlight gaps in the existing knowledge, and how this analysis fills this gap.
5. Methodology section: The research methods and the data sources used for the analysis are explained in the methodology section.
6. Results section: In the result section, the new and meaningful findings are presented with the help of descriptive statistics and illustrative graphics.
7. Discussion section: In the discussion section, data scientists should craft their main arguments in a compelling way based on the results.
8. Conclusion section: In this section, authors generalize actionable insights and make compelling statements to promote their findings. They also mention future possible deployment and applications of the research findings.
9. References: List all the references used for the analysis in the reference section.
10. Acknowledgements: In the acknowledge section, acknowledge the support of those who have enabled the work. At last, add appendices (if needed).