Guru Tegh Bahadur Institute of Technology, New Delhi

Foundations of Data Science

Course Name: B.Tech (AIML)   **Semester: III**   **SUB CODE: AIML203**

Assignment: 1

Q1. Why do we need Python for Data Science? Explain its importance depicting various features of python useful for data science.

Q2. Explain the term "Domain Knowledge" in context of Data science and python.

Q3. Differentiate between the following:

 a) Structured and Unstructured data
 b) Data Scientist, Data Analyst and Data Engineer

Q4. Explain the importance of data preprocessing and data cleaning in data science. What are the different ways of handling missing values in python?

Q5. Write a code in python for the following:

 a) Find the sum of series upto n terms (where the value of n is taken from the user)
 b) Display all prime numbers in a given range
 c) Create a calculator in python
 d) Count the number of digits in a number

Assignment: 2

Q1. Summarize the Data leaning life cycle with the help of a real-life case study.

Q2. Explain imputation. Demonstrate any two imputation techniques with the help of an example.

Q3. What is a delimiter? Highlight the steps taken to import the data (CSV) from a given folder.

Q4. Differentiate between bitwise left shift and bitwise right shift operator with the help of an example for a negative number.

Q5. How does data visualization enhance the initial exploration of a dataset? Can you provide a specific example illustrating its significance in uncovering insights from the data?

Q6. Compare and contrast the efficiency of calculating factorial of number in Python using both recursive and iterative approaches

Assignment: 3

Q1. Describe in detail about Exploratory Data Analysis?

Q2. Explain essential python libraries required for data processing and data visualization in python. Illustrate the usage with the help of a case study.

Q3. List various types of graphs/charts available in the Pyplot of matplotlib library for data visualization. Explain any two of them in brief.

Q4. What are the advantages of using Seaborn over Matplotlib for certain types of visualizations?

Q5. What is a heat map, and how can you generate one using Seaborn?

Assignment: 4

Q1. Explain the differences between supervised and unsupervised machine learning algorithms.

Q2. Explain recommender system. State some of its applications. Illustrate recommender system with the help of real-life case study.

Q3. What are the various mathematical and scientific applications for data analysis. Perform data analysis on any one of the listed applications to get a useful insight of data.

Q4. How can you visualize trends in a dataset using Python? Provide examples of relevant Python libraries for the same. Illustrate the concept with a real-life case study.

Q5. What do you mean by predictive analysis? Explain various steps involved in predictive analysis. What are the applications of predictive analysis? Illustrate predictive analysis for health care application