

Mid-Term Exam

|  |  |
| --- | --- |
| **Course Identification** | |
|  |  |
| Name of programs – Codes: | COMPUTER SCIENCE TECHNOLOGY – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING– LEA.DQ |
| Course title: | **ADVANCED DATA MANAGEMENT** |
| Course number: | 420-A15-AS |
| Group: | 7736 |
| Teacher’s name: | Kaveh Bakhtiyari |
| Duration: | EXTENDED |
| Semester: | Summer 2024 |
| **Student Identification** | |
| Name: YASHRITH CHITTOOR HARI KRISHNA Student number: 23306555  Date: **10 / 06 / 2024** Result: \_\_\_\_\_\_\_\_\_\_\_\_\_\_  I declare that this is an original work, and that I credited all content sources of which I am not the author (online and printed, images, graphics, films, etc.), in the required quotation and citation style for this work. | |
| **Standard of the Evaluated Competency** | |
|  | |

**Statement of the evaluated competency – Code**

Program Development – KP54

**Evaluated elements of the competency**

5. Evaluate the solution

6. Carryout the plan

|  |
| --- |
| **Instructions** |
|  |
| The teacher will not answer questions during the exam.You must not share the exam questions and answers during and after the exam.It is the teacher’s responsibility to identify language errors. If such errors are found, teachers may apply a penalty of up to 5% of the grade (IPEL – Article 5.7).Plagiarism, attempts at plagiarism or complicity in plagiarism during a summative evaluation results in a mark of zero (0). In the case of recidivism, in the same course or in another course, the student will be given a grade of '0' for the course in question.(IPEL – Article 5.16).Please write clearly. |
|  |
| **Mark Breakdown** |
|  |
| This evaluation is on 40 points, distributed as follows:  **TOTAL: 40 POINTS** |
|  |

**INSTRUCTION**

Answers must be provided in a Jupyer Notebook file, and all the related files should be compressed in a .zip file and attached together.

Make sure to use Markdown cells or comments in Jupyter Notebook to explain your code.

Each question must be answered clearly in that notebook.

**DATASET**

FILE: iris.csv

This dataset contains the features of the different flowers (Setosa, Versicolor, and Virginica)

**Question 1: (5 points)**

Identify the number of records (rows) and columns and list the data types of each column.

**Questions 2: (5 points)**

Get the descriptive analysis of the dataset on the numerical columns.

Write a few sentences about analysis of the dataset.

**Questions 3: (10 points)**

Plot the distributions of “Sepal Length”, and “Petal Length”.

**Questions 4: (5 points)**

Visualize the outliers in plots for numerical fields and remove the outliers.

**Questions 5: (5 points)**

Check for the missing data and handle them. You must also explain your logic on how to handle the missing values

**Questions 6: (5 points)**

Identify the categorical data and encode them properly for machine learning

**Questions 7: (5 points)**

Calculate the Pearson correlation among all the numeric fields. Then apply heatmap coloring on the resulting table to identify high and low correlations using colors.