



**Daffodil International University**  
**Department of Computer Science and Engineering**

**Faculty of Science & Information Technology**

**Midterm Examination, Fall 2021 @ DIU Blended Learning Center**

**Course Code: CSE450 (Day), Course Title: Datamining**

**Level: 4 Term: 2 Section: All**

**Instructor: FZA Modality: Open Book Exam**

**Date: Thursday 18 November, 2021 Time: 01:30 pm - 04:00 pm**

**Four hours (2:30) to support online open/case study based assessment Marks: 25**

**Directions:**

- **Students need to go through the CASE STUDY shown in this exam paper.**
- **Analyze and answer specific section based on your own thinking and work.**
- **Do not share as this will be treated as plagiarism by Blended Learning Center.**

<b>Case Study-1: Data mining basic</b>		<b>CLO1</b>
Epidemic trend analysis, timeline progression, prediction, and recommendation are critical for initiating effective public health control strategies, and data analytics play an important role in epidemiology, diagnostic, and clinical fronts. It can be a good analysis which includes the post COVID-19 in terms of effect on public health and socio economic conditions, data processing and knowledge extraction. Data sources, storage and platforms can be discussed along with discussions on data models, their performance, different big data techniques, tools and technologies. It is also the challenges in applying analytics to post pandemic scenarios, case studies and control strategies.		
<b>Q.1.a</b>	Is it possible to mitigate the post Corona challenges by using data analysis methodology? Explain your opinion	<b>5</b>
<b>Q.1.b</b>	A data collection and preprocessing technique is important for analyzing the post corona effects on public health and socio-economic scenarios. Write a brief discussion on attributes and steps of pre-processing on post corona scenarios.	<b>5</b>

<b>Case Study-2: Data mining algorithms</b>		<b>CLO2</b>
The following table presents a sample dataset where you need to predict the transportation mode of a person based on his/her car ownership along with the travel cost and income level.		

**Q.2.a**

Your task is to analyze this dataset (labeled), build a model by Calculating entropy and Gain in every step and evaluate it using the unlabeled data. Also explain what you understand by the term 'Entropy' while building a decision tree.

**10**

ID	Gender	Car Ownership	Travel Cost	Income Level	Transportation Mode
1	Male	0	Cheap	Low	Bus
2	Male	1	Cheap	Medium	Bus
3	Female	1	Cheap	Medium	Train
4	Female	0	Cheap	Low	Bus
5	Male	1	Cheap	Medium	Bus
6	Male	0	Standard	Medium	Train
7	Female	1	Standard	Medium	Train
8	Female	1	Expensive	High	Car
9	Male	2	Expensive	Medium	Car
10	Female	2	Expensive	High	Car
11	Male	1	Standard	Medium	?
12	Female	0	Cheap	Medium	?

**Q.2.b**

How would a naïve Bayes algorithm classify the given test cases? Explain in brief.

**5**