ourdiu.com | Shared by Abir Mia -231

Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Examination, Fall 2025
Course Code: CSE328, Course Title: Introduction to Data Science
Level: 3 Term: 2 Batch: 64

Time: 02:00 Hrs **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

Scenario 1: Two data scientists aim to predict whether a mango will have Good Sales or Poor Sales based on its physical and qualitative attributes. Each mango is described by a set of features collected at harvest. As a Data Scientist, your Task is to help them.

Sweetness (in 1000s) ↗	Skin Defects (in %) ↘	Sales
7.2	3	Good
5.9	12	Poor
8.5	2	Good
15.11	42	Poor
7.9	5	Good
6.1	10	Poor
8.1	4	Good

1.	**Apply** the necessary data preprocessing techniques to transform the given mango dataset into a clean, consistent, and analysis-ready form and select suitable rows for further investigation.	[10]	CO1
2.	**Find** the value of Skin Defects(%) given the Sweetness(1000s) is 5.2 for a new mango based on the preprocessed dataset. **Use** this value of Skin Defects(%) to calculate if the mango will have Good Sales or Poor Sales considering the threshold of 0.5 (Poor < 0.5 <= Good)	[10]	CO2
3.	Based on the preprocessed dataset we observe the frequency of Good Sales vs Poor Sales for the samples. Now **determine** the probability of a mango having Good Sales for the given samples. Also, **calculate** the probability that if a new batch of 5 mangoes is harvested exactly 3 of them will have Good Sales.	[10]	CO3

Scanned with CamScanner

ourdiu.com | Shared by Abir Mia -231

Skin defects in a shipment of mangoes follow a normal distribution with mean $\mu = 12\%$ and standard deviation $\sigma = 4\%$. Management is considering rejecting mangoes with more than **20 percent defects**.			
4.	**Analyze** whether the threshold of **20 percent skin defects** is a reasonable rejection rule by using the basic properties of the normal distribution. Explain your reasoning without using any statistical tables.	**10**	**CO3**

ourdiu.com

Scanned with CamScanner