United International **University (UIU)**

Dept. of Computer Science & Engineering (CSE)

CT-2

DS 1101: Fundamentals of Data Science

Total Marks: **20**   Duration: 40 Minutes

1. Consider the following data consisting of four features (Family, Height, Sporty, and Residence) and class label Potential Basketball Player. Your task is to learn a Decision Tree based on this data to predict whether a particular person is a potential basketball player. You only need to **build the tree up to two levels** (that is, the root node and another one after that). You can consider the decision of a leaf to be the class that gets the majority votes.   **[12]**

Family	Height	Sporty	Residence	Potential Basketball Player?	
Poor	Tall	Yes	City	Yes	১
Rich	Tall	No	Village	No	২
Rich	Tall	No	City	No	৩
Poor	Avg	No	Village	No	৪
Poor	Avg	No	City	No	৫
Rich	Avg	Yes	City	Yes	৬
Rich	Short	Yes	City	Yes	৭
Poor	Short	Yes	Village	No	৮
Poor	Tall	Yes	Village	Yes	৯
Poor	Tall	No	City	No	১০

2. Consider the following code:

```
import pandas as pd

import numpy as np

data = pd.DataFrame([[5., 7.5, 4., 2.5, 1., 9.],

        [1., np.nan, np.nan, 6., 4.5, 9.],

        [np.nan, np.nan, 2., 5., 7.5, np.nan],

        [np.nan, 6.5, 3., 7., 3., 5.]])
```

[Questions on the next page]

a. What will the resulting DataFrame look like? Draw as a table. [2]
b. What will the output DataFrame look like when the following codes are run separately on it? For each question below, redraw the table.

    i.   data.dropna() [2]

    ii.   data.fillna(method='ffill') [2]

    iii.   data.fillna(method='bfill') [2]