TAYLOR'S UNIVERSITY

Wisdom • Integrity • Excellence

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING DEGREE PROGRAMMES

TEST

MODULE NAME	**: Data Mining**
MODULE CODE	**: ITS61504**
DATE	**: 28th January 2022**
TIME	**: 10:00 to 18:00**
DURATION	**: 8 Hours**

Instruction to Candidates:

1. Test will be conducted on TiMES platform.
2. This paper consists of ONLY one section with ONE (1) structure question.
3. Answer ALL questions
4. Do not include the question paper in your submission

Learning Outcomes:

Develop practical data mining skills using data mining tools/languages.

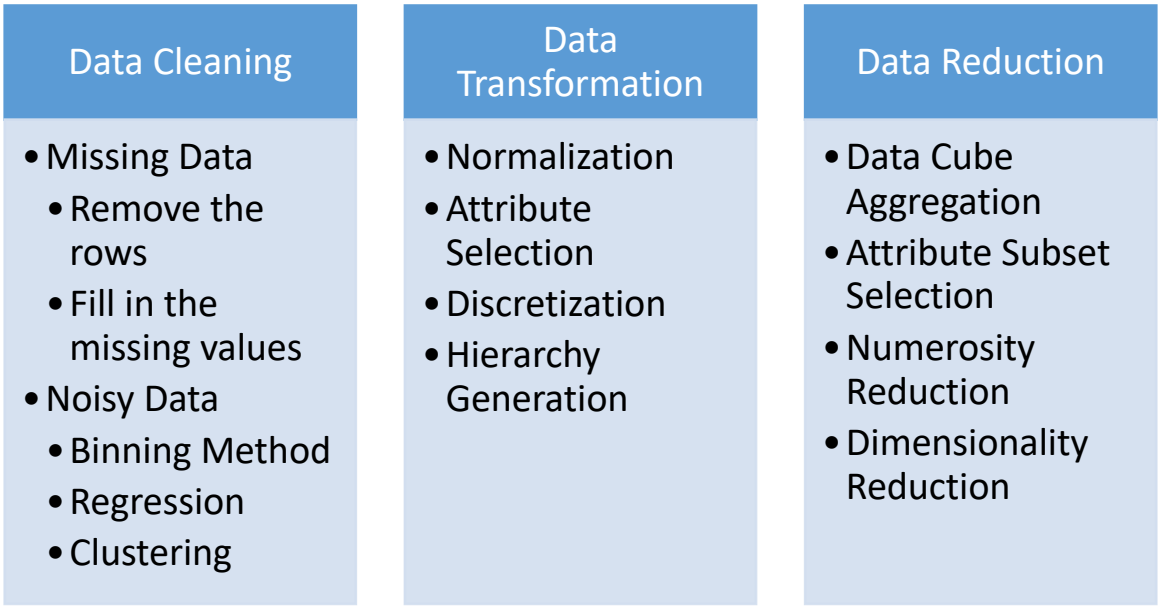
Marks Breakdown:

QUESTION	MARKS
Section A - 1 structure question	3 X 10 = 30 Marks
Marks obtained	x
Total	**x/3 Marks (10%)**

SECTION A

Answer **ALL** Questions

Data preprocessing is a data mining technique that is used to transform the raw data in a useful and efficient format. The preprocessing process can be divided into 3 major parts namely data Cleaning, Data Transformation and Data Reduction. The Figure 1 below presents the techniques that can be used to preprocess the raw data.



Preprocess the boston dataset using Jupyther notebook. The steps used should be given with good explanation and justification. Submit the ipynb file, pdf file of the ipynb file and the turnitin report. Load the dataset using the codes given below:

```
from sklearn import datasets
dir(datasets)

import pandas as pd
data = pd.DataFrame(datasets.load_boston().data)
data.columns = datasets.load_boston().feature_names
data.head(5)
```

Marking Rubric

Criteria	Weightage	Outstanding (8-10)	Mastering (5-7)	Developing (3-4)	Beginning (0-2)
Data Cleaning	10	Able to clean the dataset using the techniques learnt to handle missing data and noisy data with good explanation.	Able to clean the data using appropriate techniques.	Able to use the codes to clean the missing and noisy data.	Load the dataset as a Dataframe

Criteria	Weightage	Outstanding (8-10)	Mastering (5-7)	Developing (3-4)	Beginning (0-2)
Data Transformation	10	Able to transform the data with good explanation with supportive arguments.	Able to transform the data with explanation.	Able to transform the data.	Try to use the codes to transform the data.

Criteria	Weightage	Outstanding (8-10)	Mastering (5-7)	Developing (3-4)	Beginning (0-2)
Data Reduction	10	Able to reduce the data with good explanation with supportive arguments.	Able to reduce the data with explanation.	Able to reduce the data.	Try to use the codes to reduce the data.

End of the Paper.