TAYLOR'S UNIVERSITY

Wisdom • Integrity • Excellence

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING DEGREE PROGRAMMES

PRACTICAL

MODULE NAME	**: Data Mining**
MODULE CODE	**: ITS61504**
DATE	**: 11th 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:

Analyze using appropriate data mining techniques to achieve different purposes using various types of data set.

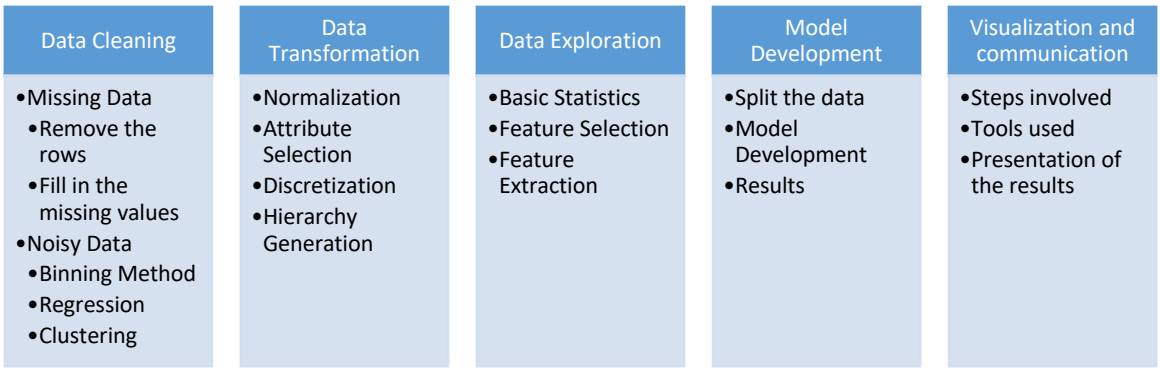
Marks Breakdown:

QUESTION	MARKS
Section A - 1 structure question	5 X 10 = 50 Marks
Marks obtained	x
Total	**x/5 Marks (10%)**

SECTION A

Answer **ALL** Questions

Data mining (knowledge discovery from data) means extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data. Data Mining consists of various steps. In this practical, do the following steps.



Analise the breast cancer 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_breast_cancer().data)
data.columns = datasets.load_breast_cancer().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 Exploration	10	An insightful and correct analysis. The following components are given. i. Basic Statistics ii. Feature Selection iii. Feature Extraction	A correct analysis that could be more complete and is not very insightful. One of the following components is missing. i. Basic Statistics ii. Feature Selection iii. Feature Extraction	An incomplete or somewhat incorrect analysis. Two of the following components are missing. i. Basic Statistics ii. Feature Selection iii. Feature Extraction	An incorrect analysis. One of the following components are given. i. Basic Statistics ii. Feature Selection iii. Feature Extraction

Criteria	Weightage	Outstanding (8-10)	Mastering (5-7)	Developing (3-4)	Beginning (0-2)
Model Development	10	An insightful and correct analysis. The following components are given. **i.** Split the data into training and testing **ii.** Develop the model using appropriate machine learning algorithm iii. Results of the prediction	A correct analysis that could be more complete and is not very insightful. One of the following components is missing. **i.** Split the data into training and testing **ii.** Develop the model using appropriate machine learning algorithm iii. Results of the prediction	An incomplete or somewhat incorrect analysis. Two of the following components are missing. **i.** Split the data into training and testing **ii.** Develop the model using appropriate machine learning algorithm iii. Results of the prediction	An incorrect analysis. One of the following components are given. **i.** Split the data into training and testing **ii.** Develop the model using appropriate machine learning algorithm iii. Results of the prediction

Criteria	Weightage	Outstanding (8-10)	Mastering (5-7)	Developing (3-4)	Beginning (0-2)
Visualization and communication	10	An insightful and correct analysis. The following components are given. i. Steps involved ii. Tools used iii. Presentation of the results	A correct analysis that could be more complete and is not very insightful. One of the following components is missing. i. Steps involved ii. Tools used iii. Presentation of the results	An incomplete or somewhat incorrect analysis. Two of the following components are missing. i. Steps involved ii. Tools used iii. Presentation of the results	An incorrect analysis. One of the following components are given. i. Steps involved ii. Tools used iii. Presentation of the results

End of the Paper.