TDS 3301 DATA MINING

Group Project

INSTRUCTIONS TO STUDENTS:

1. This assignment carries **30%**. Choose ONE (1) question only.

2. This project is a group project with a maximum of **3** members in a group.

3. If plagiarism is detected, the assignment is granted 0%.

4. The maximum number of pages is **15**, including references, using the template given. References and citations must be using APA format only.

5. Deadline for submission is on **5/1/2021**, **12pm**. Submission is to be made via Google Classroom. Timestamp will be logged.

6. You should use external dataset whenever possible.

7. Your work must consist of association rule mining, minimum two classification techniques and one clustering technique. You must provide visualization to your findings and analyze them accordingly.

8. Project leader should submit THREE items in a ZIP file: (i) a report in PDF, and (iii) LaTeX source code for preparation of the report, and (iii) a Jupyter Lab file. Name your zip file `<ID>_<Project Leader Name>.ZIP`

9. Presentation will be done in Week 12 during lecture and tutorial hours.

10. Your report should consist of the following items:

Item	**Marks**
Exploratory Data Analysis and Data Pre-Processing. Examples of question, *but not limited to*: - Can I add extra data point? - Can I do feature engineering? - How to visualize the content? - Do I need to preprocess the dataset? - Do I need to perform data imbalance treatment?	5
Feature Selection. Examples of question, *but not limited to*: - Justify why that feature selection technique(s) selected? - What is the optimal feature set?	5
Appropriate Machine Learning Techniques Used. Examples of question, *but not limited to*: - Discuss on the suitability of the techniques to the data - How do you obtain the optimal accuracy?	5
Results and Discussion. Examples of question, *but not limited to*: - Exploratory analysis - Modeling - Accuracy measurement - Findings - …etc	10
Deployment. Examples of question, *but not limited to*: - Visualization - Web Services (e.g. hosted at Heroku.com)	5

QUESTION 1: Profiling Customers in a Self-Service Coin Laundry Shop

Figures above show customers visiting a self-service laundry shop. Customers with different attires visited the shop. As a data scientist, you are expected to provide insights to the owner. You have the freedom to provide any useful insights.

[Dataset is given for QUESTION 1]

QUESTION 2: Insurance Product Recommendation

As a data scientist, you are required to develop an analytics solution to provide automation in insurance plan recommendation. The dataset consists of variables that describe the purchase history and needs by the customers of all types. Study the dataset properly and provide insights through analytics solution.

[Dataset is given for QUESTION 2]

QUESTION 3: Intelligent Decision-Making for Loan Application

Giving a score to a customer and subsequently deciding to approve or reject her loan application can be very taxing. There are many parameters to be considered. As a data scientist, you are required to support the bank to firstly provide exploratory analysis about its data and then a machine learning algorithm that can speed up loan application decision-making phase. Study the data given properly.

[Dataset is given for QUESTION 3]