Project 2

Data Analytics

a) You should submit your report in a zipped folder via LMS Drop Box folder latest by **14 June 2024**.

b) Note that plagiarism and copying are serious offences and students caught doing so will be severely punished. Late reports will also be penalized.

Project Scope:

An injection moulding machine quality experimental data are collected under three different manufacturing conditions: normal condition, condition n_1 and condition n_2. You need to complete the following tasks:

1. Open the data set that is belonged to your group
   - Understand of your data:
     - In total, how many records, how many parameters?
     - Is this a classification or regression issue?
     - How many different classes in the data?
     - How many records for each condition?
     - How many null values for each condition under each parameter?
   - Do a data pre-processing.
     - Provide the detail of steps of data checking, data pre-processing.
     - Compare of data sets before and after the processing.

2. Do statistical analysis of your data:
   - Use the method to calculate ex: mean/ median, range, std for each parameter in each condition. Analysis of the calculation result.

3. Data visualization: using the graphic method to view the data and present the finding
   You could select the chart type by yourself and may not limited to the following tasks:
   - Histogram and line chart on each column.
   - Observe which columns are the important parameters that affect the product quality using scatter plot or pair plot or heat map plot.
   - Ex: Pair plots (it is hard to see if pair plots on too many columns at same time. You could do the scatter/pair plots on important parameters or pair plots on different of parameters). Using the pair plots to find the parameters are highly correlated.

4. Data modelling and error analysis, model testing
   - Separate into training and testing data sets.
   - Build data models for this classification issues.
   - Test the model accuracy.
   - Describe your understanding of the data, model, and its accuracy.

5. Conclusion (You could do your own format and may no need to follow the below)
    - Summarize the data analytics tasks and findings for molding machine quality experimental data; Possible discussion on the influence of missing data and model limitations; Give suggestions on solutions or future research.
    - What you learnt from this course and usefulness for your IWSP, Capstone and future career.

Project files submission

Each team must submit a zip file containing:

• A formal less than 20-page report saved as a .doc file (no code should appear in the formal report).

• Your source code saved as one python file (.ipynb) + data file(.csv). Please run your code and showing the run results in python file.

• A presentation file in .ppt.

The zip file name must follow the format: Project_<Program name>_<team name>.

Any files submitted after the deadline are considered late submission. There will be a 20% penalty for late submission within 24 hours after the deadline. Submission later for more than 24 hours will be considered no submission and receives zero mark.