



School of Computer Sciences

Academic Session 2023/2024
Semester II

CMT426/CMM426–Business Intelligence and Analytics

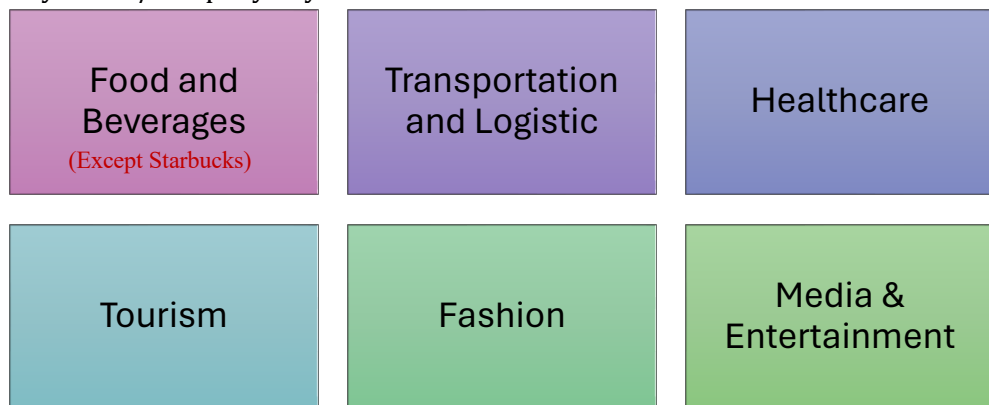
Assignment 1

CO-PO MAPPING

CL02 (PO2, P4): Construct models relevant to business intelligence and analytics using appropriate technology and software.

Perform these activities:

1. Choose any brand/company of your choice from the domain listed:



Sample Source (but not limited to):

<https://brandirectory.com/rankings/malaysia/table>

<https://business.yougov.com/content/43565-yougov-recommend-rankings-2022-malaysia>

2. Assume that you are a team of data analysts of the brand from the domain selected above. You are tasked to solve at least **two** business problem(s) **using the available dataset** from the domain selected above. You can find the dataset here (but not limited to):

*Example of brand: You choose Starbucks from Food and Beverages domain. NOTE: Given this example, you are **NOT ALLOWED** to choose Starbucks as your selected dataset for this assignment as it is a sample for benchmarking.*

- a) Kaggle: <https://www.kaggle.com/datasets>
- b) Malaysia's Open Data Portal: <https://data.gov.my/data-catalogue>
- c) KDnuggets: <https://www.kdnuggets.com/datasets/index.html>
- d) Google's Datasets Search Engine : <https://datasetsearch.research.google.com/>
- e) You may collect your own data

- f) Other reference 1: <https://towardsdatascience.com/9-best-places-to-find-machine-learning-datasets-dfdb8af5220>
 - g) Other reference 2: <https://www.v7labs.com/blog/best-free-datasets-for-machine-learning>
3. Include an introduction to the brand/company selected, the dataset used, business problems that you are trying to solve (motivation), and the objectives of the analysis (e.g., to determine, to predict).
 4. Related work – Describe the related work or state-of-the-art of business intelligence techniques that can be used to solve the stated business problem by performing a literature search and review of the previous works or studies. The purpose of this task is to understand what has been done by researchers and understand the different methods and techniques available or used in earlier research. You must include:
 - a) Review of the related works
 - b) Critically evaluated any gaps or inconsistencies in the reviews.
 - c) Summarizes the key findings of the review.
 - d) All cited works (journal paper) should be less than 5 years old.
 5. Data Exploration - Provide an overview of the dataset, e.g., how or in what context was it obtained, dataset size, what are the features/attributes, etc. Extract the relevant information to solve problems.
 6. Descriptive analytics and Data Visualisation – Include summaries of key variables, such as mean, median, mode, standard deviation, etc. Creation of appropriate dashboard, plots and charts (e.g., histograms, scatter plots, box plots, word cloud) to explore relationships between variables, trends and patterns.
 7. Data analysis – Using **Rapid Miner**, perform analysis (minimum two analysis) of the data that you have collected. The following example illustrates potential analyses (but not restricted to):
 - a) Sentiment Analysis – determine whether the sentiment leans towards positivity or negativity.
 - b) Text Classification e.g., Topic Modelling, Naïve Bayes Text Classification
 - c) Predictive modelling e.g. Decision Tree, SVM etc
 - d) Any other analysis of your preference
 8. Discussion and Conclusion - Based on your understanding, explain how the analysis performed able to solve the identified problems and improve business decision-making in the organization. You must explain the implications and justify the predictions from the results. You may also use literature to support your decision.

Tips: When determining a predictive question, the rule of thumb is to base it on what you want to do with the answer. Learn the patterns that exist in your historical data, then use those patterns to predict future outcomes.

Format:

You gain an advantage by (minimally):

1. Making your paper well-organised/balanced and easy to read.
2. Format- Font size: 12. Spacing: 1.15
3. Cover page: Course code and course title, Title, Full name of group members and metric numbers, and Submission date.
4. File format: **PDF**
5. Please name your files in this order:
 - <coursecode_Assignment1_group number e.g. G1>
6. References: List down all the references from books, journals, proceedings, websites, manual. Please cite your references. Style: APA, MLA, IEEE
7. Page limit: Not more than 15 pages
8. To adhere to the USM plagiarism policy, your report will be placed in Turn-It-In for plagiarism checking.



Submission:

Online submission: 1 June 2024, Saturday, before 11:59 p.m.

Late submission will be penalized.

Read Me: Policy

- All assignments and lab exercises **MUST** be submitted **before or on** the specified date. Late submissions without any reasons and without permission from the lecturer(s) will not be accepted. The grade for late submission (even with permission) will be reduced as determined by the lecturer(s).
- Students who copied or **plagiarized** other's work or let their work be copied or plagiarized will be given F grade for the work, test or the whole coursework component as determined by the lecturer(s). The said student may be barred from sitting for final exam and reported to the university's disciplinary board.
- **Plagiarism** (using other people's ideas and text without proper acknowledgment and using them as your own) is a serious academic offence. Offenders would be awarded grade F.

..the end

GOOD LUCK JJJ
SELAMAT MAJU JAYA