FAKULTI SAINS KOMPUTER DAN TEKNOLOGI MAKLUMAT

TEST 1
FIRST SEMESTER 2024/2025

COURSE : DATA MINING (*PERLOMBONGAN DATA*)

CODE : CSC4600

PROGRAM : _____

DATE : 17/11/2024

TIME : 9.00 PM – 10.00 PM **DURATION: 1 HOUR**

INSTRUCTIONS TO CANDIDATES:

1. There are **SIX (6)** questions in the quiz. Answer **ALL** questions.
2. Write your answers in this file.
3. The full marks for this test is **40** (10%).

Name :	CABRERA François		
Matric No. :	ES05174	**Program :**	
Group No. :	4	**Signature:**	François

This manuscript is consist of _7__ pages including the cover.

SOALAN 1/*QUESTION 1*

Terangkan secara ringkas mengenai data dan solusi perlombongan data yang dipilih oleh pasukan anda untuk Tugasan 1. [C4, 6 marks].
Briefly explain the data and data mining solutions chosen by your team in Assignment 1. [C4, 6 marks]

Jawapan/Answer:

Our project aims to reduce food waste at Carrefour by combining demand prediction and seasonal trend analysis techniques. The data used includes transactions, stock levels and weather conditions, integrated in a consistent way using unique identifiers, but also perishable data for food - we have found a table on the internet with several pieces of information on this. Data problems, such as missing values or biases, are dealt with through rigorous cleansing and normalization processes. An intuitive visual interface (tables, interactive graphs) enables you to monitor sales in real time, optimize stock management and effectively reduce losses. What's more, it's always up to date and never forgets any information.

SOALAN 2/*QUESTION* 2

Huraikan DUA (2) potensi tugas perlombongan data yang anda fokuskan dalam mengatasi cabaran keterjaminan makanan. [C3, 8 markah]
Recall the Assignments in this course that you have completed. Describe TWO (2) potential data mining tasks that you focus on in addressing food security challenges. [C3, 8 marks]

Jawapan/Answer:

For classification purposes, we have chosen
Segmentation of customers according to their purchasing habits
Classify customers according to their food preferences, such as organic, gluten-free or budget products. This makes it possible to personalize marketing campaigns.

Example Categorizing regions according to food waste
Identify regions where food waste is high to adjust stocks or raise customer awareness through anti-waste promotions.

In terms of Prediction, we have taken the following as examples
Example Anticipating weather-related demand peaks
Predict an increase in demand for cold drinks and ice cream during heat waves, or for soups and hot dishes during cold spells.

I've taken for example Forecasting out-of-stocks during promotions
Anticipate the quantities needed during a special promotion to avoid stock-outs and maximize sales.

SOALAN 3/*QUESTION 3*

Terangkan DUA (2) kriteria yang digunakan oleh pasukan anda semasa memilih set data untuk Tugasan 1. Kemudian, terangkan DUA (2) strategi yang anda gunakan untuk menilai kualiti data set data tersebut. [C3, 8 markah]
Explain TWO (2) criteria your team used when choosing the dataset for Assignment 1. Then, explain TWO (2) strategies you used to evaluate the data quality of that dataset. [C3, 8 marks]

Jawapan/Answer:

We used two criteria for selecting the dataset:

Relevance to the project objectives
The dataset was chosen based on its ability to meet the project's objectives. For example, if the goal was to analyze inventory management, the data needed to include essential fields such as stock levels, replenishment lead times, and sales volumes by product.

Data availability and accessibility
The team prioritized datasets that were complete, easy to access, and free from major constraints. This included datasets with clear documentation, a structure compatible with common tools like excel , and compliant with data privacy and sharing standards.

Two data evaluation strategies:

Checking completeness
The team assessed the dataset to detect missing values, incomplete records, or structural gaps. For instance, Python libraries such as Pandas were used to calculate the proportion of missing data in each column and identify patterns, such as product categories frequently left blank or sales records with unspecified quantities.

Validating consistency and accuracy
The data was reviewed to ensure internal consistency and accuracy. For example Delivery dates were checked to confirm they matched the operating schedules of warehouses.Product prices were analyzed to eliminate anomalies, such as excessively low or out-of-range values.Supplier names or codes were cross-checked against official directories to avoid identification errors. These steps ensure a reliable and well-structured dataset, leading to more precise and actionable results.

SOALAN 4/QUESTION 4

Nyatakan apakah data tambahan yang boleh anda gunakan bagi menjadikan solusi perlombongan anda dapat menyelesaikan masalah keterjaminan makanan? Terangkan mengapa ia penting dan bagaimana anda bercadang menggunakannya. [C4, 6 marks]
State what additional data you can use to make your mining solution able to solve food security problems? Explain why it is important and how you plan to use it. [C4, 6 marks]

Jawapan/Answer:

Fir additional data we have weather and climate data Importance: These data can be used to analyze the impact of weather conditions on demand for certain food products. For example, high temperatures can increase demand for drinks or fruit. How to use it? Combine it with sales data to perform time series analysis and regression analysis to predict demand based on weather changes. Customer profile data For Why it's important: it provides demographic information such as age, gender, income and location, which helps to understand customers' buying habits. How can I use it? Use this data for customer segmentation using clustering techniques, and to improve product recommendations using recommendation systems. Why it's important: information on competitors' prices and promotions helps you plan competitive strategies. It is used by integrating this data into association rule mining to identify products that should be promoted or repriced.

SOALAN 5/QUESTION 5

Huraikan bagaimana langkah-langkah di dalam perlombongan data yang akan anda laksanakan bagi menghasilkan solusi yang telah anda nyatakan di dalam Tugasan 1 berbeza dengan kaedah perlombongan teks. [C4, 6 markah]
Describe how the steps in data mining that you will implement to produce the solution that you have stated in Assignment 1 are different from the text mining method. [C4, 6 marks]

Jawapan/Answer:

The differences between the data mining and text mining stages are the different data Selection here are my different explanations for data mining it focuses on numerical or categorical data, such as sales figures or prices. whereas text mining works with unstructured textual data, such as customer reviews or product descriptions.Data pre-processing, as described above for data mining, includes steps such as data normalization, missing value management and attribute combination. For text mining involves natural language processing techniques. For analysis methods, data mining aims to find patterns or relationships using techniques such as regression, clustering or classification.have therefore for Text Mining The analysis focusing on tasks such as sentiment analysis, document clustering or entity recognition.

SOALAN 6/QUESTION 6

Cadangkan DUA (2) soalan yang boleh digunakan oleh mana-mana penilai untuk menilai penyelesaian perlombongan data yang akan anda bangunkan setelah melengkapkan kursus ini, dan nyatakan mengapa soalan ini penting [C4, 6 markah].
Propose TWO (2) questions that any assessor can use to assess the data mining solution you will develop upon completing this course, and state why these questions are important [C4, 6 marks].

Jawapan/Answer:

For my various questions, I thought of the following

How effectively does this solution address food safety issues?
Why this is important: this question assesses the extent to which the solution solves the main problem using the data provided.

And for the second question, I thought of this

Can the data mining model be extended to larger datasets or applied to other contexts?

Why this is important: this question assesses whether the solution can be implemented in other locations or for other organizations with similar data structures.

- END OF QUESTION -