
UNIVERSITI MALAYSIA PERLIS

Peperiksaan Akhir Semester Pertama
Sidang Akademik 2025/2026

Februari 2026

IMC11203 – Introduction to Computing & Data Science
[*Pengenalan kepada Pengkomputeran & Sains Data*]

Masa : 2 jam

Please make sure that this question paper has **EIGHT (8)** printed pages, including this front page, before you start the examination.
*[Sila pastikan kertas soalan ini mengandungi **EIGHT (8)** muka surat yang bercetak termasuk muka hadapan sebelum anda memulakan peperiksaan ini.]*

This question paper has **FOUR (4)** questions. Answer **ALL** questions. Each question carries 20 marks, contributing to a total of 80 marks.
*[Kertas soalan ini mengandungi **EMPAT (4)** soalan. Jawab **SEMUA** soalan. Setiap soalan adalah 20 markah, dengan jumlah keseluruhan sebanyak 80 markah.]*

Dicetak oleh Unit Peperiksaan & Pengijazahan, Bahagian Pengurusan Akademik, Pusat Pengurusan Akademik UniMAP

ANSWER **ALL** QUESTIONS
*[JAWAB **SEMUA** SOALAN]*

(CO1, PO1, C2)
Question 1
[Soalan 1]

(a) Describe **TWO (2)** major causes of large-scale data growth in modern computing environments and explain how each contributes to the Big Data problem.
*[Terangkan **DUA (2)** punca utama pertumbuhan data berskala besar dalam persekitaran pengkomputeran moden dan jelaskan bagaimana setiap punca menyumbang kepada masalah 'Data Raya'.]*

(4 Marks/ *Markah*)

(b) Explain how data replication improves fault tolerance in distributed storage systems.
[Terangkan bagaimana replikasi data dapat meningkatkan toleransi ralat dalam sistem storan teragih.]

(4 Marks/ *Markah*)

(c) Differentiate between task parallelism and data parallelism.
[Bezakan antara paralelisme tugas dan paralelisme data.]

(4 Marks/ *Markah*)

(d) Describe **TWO (2)** scenarios where a company would use a private cloud.
*[Huraikan **DUA (2)** senario di mana sesebuah syarikat akan menggunakan 'private cloud'.]*

(4 Marks/ *Markah*)

(e) Discuss the difference between time and space complexity.
[Bincangkan perbezaan antara kerumitan masa dan kerumitan ruang.]

(4 Marks/ *Markah*)

(CO1, PO1, C2)
Question 2
[Soalan 2]

(a) Explain the concept of the Data Science Lifecycle and describe why it is characterized as iterative and cyclical rather than a linear process. Provide **ONE (1)** example of situations where you would need to return to an earlier phase.
*[Terangkan konsep Kitar Hayat Sains Data dan huraikan mengapa ia dicirikan sebagai proses yang berulang dan bersifat kitaran dan bukannya proses linear. Berikan **SATU (1)** contoh situasi di mana anda perlu kembali semula ke fasa terdahulu.]*

(6 Marks/ *Markah*)

(b) Discuss the Exploratory Data Analysis (EDA) process.
[Bincangkan proses Analisis Data Eksploratori (EDA).]

(4 Marks/ *Markah*)

(c) The following **Table 2.1** illustrates several processes carried out by a data science team from a retail company to enhance its marketing strategy by following the Knowledge Discovery in Databases (KDD) steps to uncover useful patterns and insights. For each description, identify the corresponding KDD step (Selection, Preprocessing, Transformation, Data Mining, or Interpretation/Evaluation) in the third column. Note: The steps are not listed in sequential order.
*[**Jadual 2.1** berikut menunjukkan beberapa proses yang dilaksanakan oleh pasukan sains data daripada sebuah syarikat runcit bagi mempertingkat strategi pemasarannya dengan mengikuti langkah-langkah Penemuan Pengetahuan dalam Pangkalan Data (KDD) untuk mengenal pasti corak dan maklumat yang berguna. Bagi setiap penerangan, kenal pasti langkah KDD yang sepadan ('Selection', 'Preprocessing', 'Transformation', 'Data Mining', atau 'Interpretation/Evaluation') pada lajur ketiga. Nota: Langkah-langkah tersebut tidak disenaraikan mengikut urutan sebenar.]*

Table 2.1
[Jadual 2.1]

Process *[Proses]*	Description *[Penerangan]*	KDD Step *[Langkah KDD]*
P	The data scientists apply clustering algorithms (e.g., K-Means) to group customers based on purchasing habits. They discover three distinct customer segments, which are (i) frequent high spenders, (ii) occasional buyers and (iii) discount seekers. *[Ahli sains data menggunakan algoritma pengelompokan (contohnya, K-Means) untuk mengelompokkan pelanggan berdasarkan tabiat pembelian mereka. Mereka menemui tiga segmen pelanggan yang berbeza, iaitu: (i) pelanggan yang kerap berbelanja besar, (ii) pembeli secara berkala, dan (iii) pencari diskaun.]*	

Q	To prepare for modeling, the team selects key attributes such as customer age group, average purchase value and frequency of visits. They also convert categorical data (e.g., gender) into numerical form and normalize numeric variables. *[Bagi membuat persediaan untuk pemodelan, pasukan tersebut memilih atribut utama seperti kumpulan umur pelanggan, nilai purata pembelian, dan kekerapan lawatan. Mereka juga menukar data kategori (contohnya, jantina) kepada bentuk berangka serta menormalkan pemboleh ubah berangka.]*	
R	The marketing team examines the cluster results and interprets them: • Segment 1 → target for loyalty programs • Segment 2 → offer occasional discounts • Segment 3 → send promotional advertisement They present a report to management with actionable insights. *[Pasukan pemasaran meneliti hasil pengelompokan dan mentafsirkannya seperti berikut:* • *Segmen 1 → disasarkan untuk program kesetiaan pelanggan* • *Segmen 2 → ditawarkan diskaun berkala* • *Segmen 3 → dihantar iklan promosi* *Mereka membentangkan laporan kepada pihak pengurusan dengan pandangan serta cadangan yang boleh dilaksanakan.]*	
S	The team gathers raw data from multiple sources which are online transaction logs, customer profile database and web browsing behavior data. They combine these into a single dataset for analysis. *[Pasukan tersebut mengumpulkan data mentah daripada pelbagai sumber iaitu log transaksi dalam talian, pangkalan data profil pelanggan, dan data tingkah laku pelayaran web. Mereka menggabungkan semua data ini ke dalam satu set data tunggal untuk tujuan analisis.]*	
T	Upon examining the raw data, analysts find missing values in customer ages, duplicate purchase records, and inconsistent date formats. They remove duplicates, fill missing values, and standardize all date fields. *[Apabila meneliti data mentah tersebut, penganalisis mendapati terdapat nilai hilang bagi umur pelanggan, rekod pembelian yang berulang, serta format tarikh yang tidak konsisten. Mereka menghapuskan data pendua, mengisi nilai yang hilang, dan menyeragamkan semua format tarikh.]*	

(10 Marks/ *Markah*)

(CO2, PO2, C3)
Question 3
[Soalan 3]

(a) A company is considering investments in two IT projects as shown in **Table 3.1**.
*[Sebuah syarikat sedang mempertimbangkan pelaburan dalam dua projek IT seperti yang ditunjukkan dalam **Jadual 3.1**.]*

Table 3.1
[Jadual 3.1]

Project *[Projek]*	Investment Cost (RM) *[Kos Pelaburan (RM)]*	Expected Returns over 3 Years (RM) *[Anggaran Pulangan dalam 3 Tahun (RM)]*
Project A: Cloud Migration *[Projek A: Migrasi Awan]*	200,000	280,000
Project B: Cybersecurity Upgrade *[Projek B: Naiktaraf Keselamatan Siber]*	320,000	420,000

(i) Apply the Return on Investment (ROI) formula to calculate the ROI for both projects and interpret the results to evaluate their financial viability.
[Aplikasikan rumus Pulangan Ke Atas Pelaburan (ROI) untuk mengira ROI bagi kedua-dua projek dan interpretasikan keputusan bagi menilai daya maju kewangan projek tersebut.]

(6 Marks/ *Markah*)

(ii) Discuss the outcomes and recommendation on which project the company should prioritize or whether both investments should proceed.
[Bincangkan dapatan dan syor mengenai projek yang sepatutnya diberi keutamaan oleh syarikat atau sama ada kedua-dua pelaburan perlu diteruskan.]

(6 Marks/ *Markah*)

(b) A mobile application company collects users' location data for advertising purposes. Based on the General Data Protection Regulation (GDPR), identify **TWO (2)** relevant data protection principles and provide **TWO (2)** legal obligations (one for each principle) where the company must fulfil to ensure compliance.
*[Sebuah syarikat aplikasi mudah alih mengumpul data lokasi pengguna untuk tujuan pengiklanan. Berdasarkan Peraturan Perlindungan Data Am (GDPR), kenal pasti **DUA (2)** prinsip perlindungan data yang berkaitan dan berikan **DUA (2)** tanggungjawab undang-undang (satu untuk setiap prinsip) yang mesti dipenuhi oleh syarikat untuk memastikan pematuhan.]*

(8 Marks/ *Markah*)

(CO2, PO2, C3)
Question 4
[Soalan 4]

TechStart Inc., a rapidly growing fintech startup, has hired you as their lead developer. You notice several concerning practices:
[TechStart Inc., sebuah syarikat permulaan 'fintech' yang berkembang dengan pesat, telah melantik anda sebagai ketua pembangun mereka. Anda mendapati beberapa amalan yang membimbangkan, iaitu:]

- No code review process exists.
  [Tiada proses kajian semula kod wujud.]
- Documentation is minimal or absent.
  [Dokumentasi adalah minima atau tidak wujud sama sekali.]
- Developers frequently copy code from online sources without understanding licensing implications.
  [Pembangun sering menyalin kod daripada sumber dalam talian tanpa memahami implikasi lessen.]
- The team rushes feature without adequate testing.
  [Pasukan terburu-buru melaksanakan ciri tanpa pengujian yang mencukupi.]
- Security practices are inconsistent.
  [Amalan keselamatan adalah tidak konsisten.]

Based on this scenario, answer the following questions.
[Berdasarkan scenario ini, jawab soalan-soalan yang berikut.]

(a) Apply your understanding of professionalism in computing to identify and explain **THREE (3)** violations of the ACM Code of Ethics in the given scenario. For **each violation, explain ONE (1)** potential consequence or risk. In your answer, clearly specify the relevant ACM ethical code associated with each violation. Refer **Appendix A** for the list of ACM Code of Ethics.
*[Aplikasikan kefahaman anda tentang profesionalisme dalam bidang pengkomputeran untuk mengenal pasti dan menjelaskan **TIGA (3)** pelanggaran Kod Etika ACM dalam senario yang diberikan. Bagi **setiap pelanggaran, jelaskan SATU (1)** implikasi pelanggaran atau risiko. Dalam jawapan anda, menyatakan dengan jelas kod etika ACM yang berkaitan bagi setiap pelanggaran. Rujuk **Appendik A** bagi senarai Kod Etika ACM.]*

(12 Marks/ *Markah*)

(b) As the lead developer, discuss and elaborate **TWO (2)** measures you would implement to ensure ethical handling of code and intellectual property within TechStart Inc.
*[Sebagai ketua pembangun, bincang dan perjelaskan **DUA (2)** langkah yang anda akan ambil untuk memastikan pengurusan kod dan harta intelek secara beretika di dalam TechStart Inc.]*

(8 Marks/ *Markah*)

APPENDIX A
[LAMPIRAN A]

ACM Code of Ethics and Professional Conduct
[Kod Etika dan Tatakelola Profesional ACM]

1. GENERAL ETHICAL PRINCIPLES	1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing. 1.2 Avoid harm. 1.3 Be honest and trustworthy. 1.4 Be fair and take action not to discriminate. 1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts. 1.6 Respect privacy. 1.7 Honor confidentiality.
2. PROFESSIONAL RESPONSIBILITIES	2.1 Strive to achieve high quality in both the processes and products of professional work. 2.2 Maintain high standards of professional competence, conduct, and ethical practice. 2.3 Know and respect existing rules pertaining to professional work. 2.4 Accept and provide appropriate professional review. 2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks. 2.6 Perform work only in areas of competence. 2.7 Foster public awareness and understanding of computing, related technologies, and their consequences. 2.8 Access computing and communication resources only when authorized or when compelled by the public good. 2.9 Design and implement systems that are robustly and usably secure
3. PROFESSIONAL LEADERSHIP PRINCIPLES	3.1 Ensure that the public good is the central concern during all professional computing work. 3.2 Articulate, encourage acceptance of, and evaluate fulfilment of social responsibilities by members of the organization or group. 3.3 Manage personnel and resources to enhance the quality of working life. 3.4 Articulate, apply, and support policies and processes that reflect the principles of the Code. 3.5 Create opportunities for members of the organization or group to grow as professionals. 3.6 Use care when modifying or retiring systems. 3.7 Recognize and take special care of systems that become integrated into the infrastructure of society.
4. COMPLIANCE WITH THE CODE	4.1 Uphold, promote, and respect the principles of the Code. 4.2 Treat violations of the Code as inconsistent with membership in the ACM.

APPENDIX B
[LAMPIRAN B]

CO – PO Information
[Maklumat CO – PO]

Course Outcomes (COs)

CO1	Ability to explain key concepts of computing systems, data science methodologies, and professional practices in data science contexts.
CO2	Ability to apply key concepts of computing systems, data science methodologies, and professional practices to solve data science problems.

Programme Outcomes (POs)

PO1	Knowledge & Understanding – Analyse data, concepts, principles and theories relating to Data Science.
PO2	Cognitive Skills – Apply appropriate concepts and methods for optimised Data Science solutions.

-oooOooo-